

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2004, 09:45:20 ; Search time 120 Seconds
(without alignments)
28.255 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPLNTHNGQKL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	12	5 AAM50916	Aam50916 Tumour in
2	64	100.0	19	5 AAM50917	Aam50917 Tumour in
3	64	100.0	19	5 AAM50918	Aam50918 Tumour in
4	42	65.6	258	6 ADB10796	ADB10796 Alloiococ
5	42	65.6	804	4 ABG28800	ABg28800 Novel hum
6	42	65.6	804	4 ABG28023	ABg28023 Novel hum
7	42	65.6	807	4 ABG28665	ABg28665 Novel hum
8	42	65.6	1023	4 ABG28668	ABg28668 Novel hum
9	42	65.6	1302	4 ABG64919	ABg64919 Drosophil
10	42	65.6	1325	4 AAG98256	Aag98256 Escherich
11	40	62.5	61	4 AB341576	Abb41576 Peptide #
12	40	62.5	61	4 AAM33368	Aam33368 Peptide #
13	40	62.5	61	4 AB224416	Abm24416 Protein #
14	40	62.5	61	4 AAM72526	Aam72526 Human bon
15	40	62.5	61	4 AAM62447	Aam62447 Human bra
16	40	62.5	61	4 ABG57018	ABg57018 Human liv
17	40	62.5	61	5 ABG44904	ABg44904 Human pep
18	39	60.9	243	6 ABU49842	Abu49842 protein e
19	39	60.9	256	6 ABM67741	ABm67741 Phototrab
20	39	60.9	392	4 AB94887	AB94887 Human pro
21	39	60.9	610	5 ABG31908	ABg31908 Zinc fing
22	39	60.9	610	6 ADA55036	Ada55036 Human pro
23	39	60.9	2176	4 ABM61386	ABm61386 Drosophil
24	38	59.4	82	5 ABP04929	ABp04929 Human ORF
25	38	59.4	117	7 ADE08998	Ade08998 Novel pro

26	38	59.4	150	5	ABE53387	Abb53387 Lactococc
27	38	59.4	247	3	AAy83081	Aay83081 F-box pro
28	38	59.4	247	5	AAO22467	Aao22467 Human F-b
29	38	59.4	376	7	ADE07936	Ade07936 Novel pro
30	38	59.4	408	4	AAB48304	Aab48304 Human ZF2
31	38	59.4	414	5	ABG70045	ABg70045 Human pre
32	38	59.4	444	4	AAG62447	Aag62447 Zinc fing
33	38	59.4	444	4	AAG62620	Aag62620 Male ster
34	38	59.4	444	4	AAG64526	Aag64526 P. hybrid
35	38	59.4	444	4	AAG64874	Aag64874 Pollen fe
36	38	59.4	444	7	ADC51677	Adc51677 Garden pe
37	38	59.4	444	7	ADD02825	Add02825 Petunia h
38	38	59.4	460	6	ABJ19183	Abj19183 Pathogen
39	38	59.4	460	6	ABU42681	Abu42681 Protein e
40	38	59.4	467	5	ABP38390	ABp38390 Staphyloc
41	38	59.4	468	5	ABP43604	ABp43604 Clone MGC
42	38	59.4	471	3	AAy90287	Aay90287 Human pep
43	38	59.4	471	4	AAAB93481	Aaab93481 Human pro
44	38	59.4	471	5	AAAM4764	Aam47764 F-cassett
45	38	59.4	472	4	AAAG82404	Aag82404 S. epider

ALIGNMENTS

RESULT 1

AAM50916

ID AAM50916 standard; peptide; 12 AA.

XX

AC AAM50916;

XX

DT 07-MAY-2002 (first entry)

XX

DE Tumour infiltrating peptide HN-1.

XX

KW Tumour infiltrating peptide; HN-1; head and neck cancer; HNSCC;

KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX

OS Synthetic.

XX

PN WO200202147-A2.

XX

PD 10-JAN-2002.

XX

PF 02-JUL-2001; 2001WO-US021518.

XX

PR 30-JUN-2000; 2000US-0215491P.

XX

PA (TEXA) UNIV TEXAS SYSTEM.

PI Clayman G, Hong PD;

XX

DR WPI; 2002-195737/25.

 XX |

PT Peptide internalized by a tumor cell useful for targeted delivery of anticancer drugs.

XX

PS Claim 2; Page 71; 104pp; English.

XX
The present sequence is that of a synthetic peptide, termed HN-1, that is specifically internalised by human head and neck squamous carcinoma cells (HNSCC), or certain other solid tumour tissue cells, such as breast cancer cells. HN-1 was identified by screening a phage M13 peptide library displaying over 10 power 9 peptides. The screening method was based on the ability of HNSCC line MDAl67rv cells to uptake peptides by endocytosis at 37 degrees C. The cells exhibited an approximately 10-fold greater internalisation potential for HN-1 than normal human fibroblasts. The peptide localised in cytoplasm after entry. In vivo, i.v. injected HN-1 peptide localised to HNSCC xenograft formed in nude mice. Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic, apoptotic or DNA-damaging drug, such as taxol. The compositions are used in claimed methods for killing a tumour cell, especially an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal cavity, larynx, thyroid,

CC parathyroid, salivary gland, face or neck skin or cervical lymph node
 CC cell), breast cancer cell or other solid tumour cell. Also claimed are:
 CC methods for detecting cancer by labeling HN-1 with a detectable label;
 CC tumour detection, tumour imaging and tumour treating kits; methods for
 CC killing tumour cells in which a composition comprising an antitumour
 CC compound conjugated to HN-1 is administered with radiotherapy,
 CC chemotherapy, surgery or a gene therapy composition; and a method for the
 CC isolating of an internalising peptide by phage display library screening.
 CC The peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide is non
 CC -toxic, non-immunogenic, stable in vivo, protects its cargo during
 CC transit, and accumulates in a tumour within 48 hours
 CC
 XX
 SQ Sequence 12 AA;

Query Match 100.0%; Score 64; DB 5; Length 12;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TSPLNIHNGQKL 12
 |||||
 Db 1 TSPLNIHNGQKL 12

RESULT 2
 AAM50917
 ID AAM50917 standard; peptide; 19 AA.

XX AC AAM50917;

XX DT 07-MAY-2002 (first entry)

XX DE Tumour infiltrating peptide HN-2.

XX KW Tumour infiltrating peptide; HN-2; head and neck cancer; HNSCC;
 XX breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX OS Synthetic.

XX PH Key Location/Qualifiers
 XX FT Peptide 4. .15
 XX FT /note= "corresponds to HN-1"

XX PN WO200202147-A2.

XX PD 10-JAN-2002.

XX PF 02-JUL-2001; 2001WO-US021518.

XX PR 30-JUN-2000; 2000US-0215491P.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Clayman G, Hong FD;

XX DR WPI; 2002-195737/25.

XX PT Peptide internalized by a tumor cell useful for targeted delivery of
 XX anticancer drugs.

XX PS Example 2; Page 76; 104pp; English.

CC The present sequence is that of a synthetic peptide, termed HN-2, which
 CC is based on the claimed tumour infiltrating peptide HN-1 (see AAM50916)
 CC with additional N- and C-terminal amino acid residues. HN-1 is
 CC specifically internalised by human head and neck squamous carcinoma cells
 CC (HNSCC) and certain other solid tumour tissue cells, such as breast
 CC cancer cells. The additional amino acid residues of HN-2 did not inhibit
 CC cell internalisation of the peptide; HN-1 internalisation is position-
 CC independent but sequence-dependent. Claimed compositions comprise HN-1
 CC and a chemotherapeutic, cytotoxic, apoptotic or DNA-damaging drug, such
 CC as taxol. The compositions are used in claimed methods for killing a
 CC tumour cell, especially an HNSCC (oral cavity, pharynx, throat, paranasal

CC sinus, nasal cavity, larynx, thyroid, parathyroid, salivary gland, face
 CC or neck skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer by
 CC labeling HN-1 with a detectable label; tumour detection, tumour imaging
 CC and tumour treating kits; methods for killing tumour cells in which a
 CC composition comprising an antitumour compound conjugated to HN-1 is
 CC administered with radiotherapy, chemotherapy, surgery or a gene therapy
 CC composition; and a method for the isolating of an internalising peptide
 CC by phage display library screening. The peptide provides the necessary
 CC dose of a drug specifically to the tumours, avoiding harmful side effects
 CC on other cells. The peptide is non-toxic, non-immunogenic, stable in
 CC vivo, protects its cargo during transit, and accumulates in a tumour
 CC within 48 hours
 CC
 XX
 SQ Sequence 19 AA;

Query Match 100.0%; Score 64; DB 5; Length 19;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TSPLNIHNGQKL 12
 |||||
 Db 4 TSPLNIHNGQKL 15

RESULT 3
 AAM50918

ID AAM50918 standard; peptide; 19 AA.

XX AC AAM50918;

XX DT 07-MAY-2002 (first entry)

XX DE Tumour infiltrating peptide HN-3.

XX KW Tumour infiltrating peptide; HN-3; head and neck cancer; HNSCC;
 XX breast cancer; drug delivery; diagnosis; imaging; gene therapy.

XX OS Synthetic.

XX PH Key Location/Qualifiers
 XX FT Peptide 8. .19
 XX FT /note= "corresponds to HN-1"

XX PN WO200202147-A2.

XX PD 10-JAN-2002.

XX PF 02-JUL-2001; 2001WO-US021518.

XX PR 30-JUN-2000; 2000US-0215491P.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Clayman G, Hong FD;

XX DR WPI; 2002-195737/25.

XX PT Peptide internalized by a tumor cell useful for targeted delivery of
 XX anticancer drugs.

XX PS Example 2; Page 76; 104pp; English.

CC The present sequence is that of a synthetic peptide, termed HN-3, which
 CC is based on the claimed tumour infiltrating peptide HN-1 (see AAM50916)
 CC with additional N- terminal amino acid residues. HN-1 is specifically
 CC internalised by human head and neck squamous carcinoma cells (HNSCC) and
 CC certain other solid tumour tissue cells, such as breast cancer cells. The
 CC additional N-terminal amino acid residues of HN-3 did not inhibit cell
 CC internalisation of the peptide; HN-1 internalisation is position-
 CC independent but sequence-dependent. Claimed compositions comprise HN-1
 CC and a chemotherapeutic, cytotoxic, apoptotic or DNA-damaging drug, such
 CC as taxol. The compositions are used in claimed methods for killing a

CC tumour cell, especially an HNSCC (oral cavity, pharynx, throat, paranasal
 CC sinus, nasal cavity, larynx, thyroid, parathyroid, salivary gland, face
 CC or neck skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer by
 CC labeling HN-1 with a detectable label; tumour detection, tumour imaging
 CC and tumour treating kits; methods for killing tumour cells in which a
 CC composition comprising an antitumour compound conjugated to HN-1 is
 CC administered with radiotherapy, chemotherapy, surgery or a gene therapy
 CC composition; and a method for the isolating of an internalising peptide
 CC by phage display library screening. The peptide provides the necessary
 CC dose of a drug specifically to the tumours, avoiding harmful side effects
 CC on other cells. The peptide is non-toxic, non-immunogenic, stable in
 CC vivo, protects its cargo during transit, and accumulates in a tumour
 CC within 48 hours
 XX
 SQ Sequence 19 AA;

Query Match 100.0%; Score 64; DB 5; Length 19;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPINIHNGQKL 12
 |||||
 Db 8 TSPINIHNGQKL 19

RESULT 4
 ADB10796
 ID ADB10796 standard; protein; 258 AA.

XX AC ADB10796;

XX DT 20-NOV-2003 (first entry)

XX DE Alloicoccus otitis antigenic protein SEQ ID NO:4264.

XX KW Alloicoccus otitidis; antigenic protein; immunogenic; immunisation;
 gene therapy; Gram-positive bacterium; infection.

XX OS Alloicoccus otitis.

XX PN WO2003048304-A2.

XX PD 12-JUN-2003.

XX PF 25-NOV-2002; 2002WO-US036123.

XX PR 29-NOV-2001; 2001US-0333777P.

XX PR 18-NOV-2002; 2002US-0426742P.

XX PA (AMHP) WYETH HOLDINGS CORP.

XX PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX DR WPI; 2003-505284/47.

XX DR N-PSDB; ADB10799.

XX PT New Alloicoccus otitidis polynucleotides and polypeptides, useful for
 treating and diagnosing diseases, drug screening assays and monitoring of
 effects during drug clinical trials.

XX FS Claim 33; SEQ ID NO 4264; 1019pp; English.

XX CC The present invention describes an isolated polynucleotide (I) of
 CC Alloicoccus otitidis genomic DNA, which encodes an antigenic protein.
 CC Alloicoccus otitidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I); its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the

CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloicoccus otitidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloicoccus
 CC otitidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloicoccus
 CC otitidis. The present sequence represents an Alloicoccus otitidis
 CC antigen protein from the present invention.

XX SQ Sequence 258 AA;

Query Match 65.6%; Score 42; DB 6; Length 258;

Best Local Similarity 58.3%; Pred. No. 12;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPINIHNGQKL 12

:||| |||

Db 10 SSPLIHQGRKM 21

RESULT 5

ABG28800

ID ABG28800 standard; protein; 804 AA.

XX AC ABG28800;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #28791.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS92987.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX FS Claim 20; SEQ ID NO 59159; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 804 AA;

Query Match 65.6%; Score 42; DB 4; Length 804;
 Best Local Similarity 88.9%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGQKL 12
 |||||
 Db 325 LNIHGGQKL 333

RESULT 6
 ID ABG29023 standard; protein; 804 AA.

AC ABG29023;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #29014.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS93210.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 59382; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 804 AA;

Query Match 65.6%; Score 42; DB 4; Length 804;
 Best Local Similarity 88.9%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGQKL 12
 |||||
 Db 325 LNIHGGQKL 333

RESULT 7

ABG28665

ID ABG28665 standard; protein; 807 AA.

AC ABG28665;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #28656.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS92852.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 59024; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 907 AA;
 SQ

Query Match 65.6%; Score 42; DB 4; Length 807;
 Best Local Similarity 88.9%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LNIHNGQKL 12
 Db 328 LNIHGGQKL 336
 |||||
 |||||

RESULT 8
 ABG28668
 ID ABG28668 standard; protein; 1023 AA.
 XX
 AC ABG28668;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #28659.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 PI
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS92855.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 59027; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 1023 AA;
 SQ

Query Match 65.6%; Score 42; DB 4; Length 1023;
 Best Local Similarity 88.9%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LNIHNGQKL 12
 Db 439 LNIHGGQKL 447
 |||||
 |||||

RESULT 9
 ABB64919
 ID ABB64919 standard; protein; 1302 AA.
 XX
 AC ABB64919;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 21549.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL09022.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 21549; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB2072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ

Query Match 65.6%; Score 42; DB 4; Length 1302;
 Best Local Similarity 72.7%; Pred. No. 76;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY      1 TSPLNHNHNGOK 11
Db      673 TSPLNFEKGOK 683

RESULT 10
AAG98256
ID   AAG98256 standard; protein; 1325 AA.
XX
AC   AAG98256;
XX
DT   21-SEP-2001 (first entry)
XX
DE   Escherichia coli protein sequence SEQ ID NO:304.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
XX   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX   bacterial growth inhibition.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US034419.
XX
PR   23-DEC-1999; 99US-0173005P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA, Ohlsen KL, Zyskind JW;
XX
WPI: 2001-457376/49.
XX
N-PSDB; AAH81312.
XX
Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents.
XX
PS   Claim 19; Page 445-448; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I), consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81292 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in gene
CC   therapy. Expression of (I) in a microorganism inhibits proliferation of
CC   the microorganism, and the manufactured antibiotic is useful for reducing
CC   the activity or level of a gene product required for proliferation of a
CC   microorganism in a subject, specifically humans. The nucleic acids that
CC   inhibit bacterial growth or proliferation can be used as antisense
CC   therapeutics for killing bacteria. In addition to therapeutic
CC   applications, the nucleic acid sequences complementary to sequences
CC   required for proliferation can be used as diagnostic tools. For example,
CC   nucleic acid probes complementary to proliferation-required sequences
CC   that are specific for particular species of microorganisms can be used as
CC   probes to identify particular microorganism species in clinical
CC   specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
CC   given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
CC   oligonucleotides, which are used in the exemplification of the present
CC   invention
XX
SQ   Sequence 1325 AA;

Query Match      65.8%; Score 42; DB 4; Length 1325;
Best Local Similarity 88.9%; Pred. No. 77;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 LTNHNGOKL 12
Db      991 LNHGQKL 999

RESULT 11
AEB41576
ID   AEB41576 standard; peptide; 61 AA.
XX
AC   AEB41576;
XX
DT   04-FEB-2002 (first entry)
XX
DE   Peptide #9082 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
XX
PR   26-MAY-2000; 2000US-0207456P.
XX
PR   30-JUN-2000; 2000US-00608408.
XX
PR   03-AUG-2000; 2000US-00632366.
XX
PR   21-SEP-2000; 2000US-0234687P.
XX
PR   27-SEP-2000; 2000US-0236359P.
XX
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2001-483447/52.
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 34211; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 61 AA;

Query Match      62.5%; Score 40; DB 4; Length 61;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNHNHNGOK 11
Db      35 TSPKHKNNGOK 45

RESULT 12
AAM35368
ID   AAM35368 standard; protein; 61 AA.
XX
AC   AAM35368;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Peptide #9405 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
XX   genetic disorder.

```

```

OS Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX SQ
Query Match 62.5%; Score 40; DB 4; Length 61;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPINIHNGOK 11
DB 35 TSPRHKNKGOK 45

RESULT 13
ABB25416
ID ABB25416 standard; protein; 61 AA.
XX ABB25416;
XX 23-JAN-2002 (first entry)
XX Protein #7415 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.

OS Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 35637; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX SQ
Query Match 62.5%; Score 40; DB 4; Length 61;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPINIHNGOK 11
DB 35 TSPRHKNKGOK 45

RESULT 14
AAM75256
ID AAM75256 standard; protein; 61 AA.
XX AAM75256;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35562.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 15; SEQ ID NO 27186; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ
Query Match 62.5%; Score 40; DB 4; Length 61;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPINIHNGOK 11
DB 35 TSPRHKNKGOK 45

RESULT 14
AAM75256
ID AAM75256 standard; protein; 61 AA.
XX AAM75256;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 35562.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 35637; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX SQ
Query Match 62.5%; Score 40; DB 4; Length 61;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPINIHNGOK 11
DB 35 TSPRHKNKGOK 45

RESULT 13
ABB25416
ID ABB25416 standard; protein; 61 AA.
XX ABB25416;
XX 23-JAN-2002 (first entry)
XX Protein #7415 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.

```

PT gene expression in human bone marrow.
 XX Example 4; SEQ ID NO 35562; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention
 XX
 XX Sequence 61 AA;
 SQ

Query Match 62.5%; Score 40; DB 4; Length 61;
 Best Local Similarity 72.7%; Pred. No. 5.3;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNHNGOK 11
 |||:||||
 Db 35 TSPKHNGOK 45

RESULT 15
 AAM62447
 ID AAM62447 standard; protein; 61 AA.
 XX AC
 XX AAM62447;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34552.
 XX
 XX Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
 XX
 XX Homo sapiens.
 XX
 XX WO200157275-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000667.
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLB-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2001-483446/52.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 DR brains.
 XX
 XX Example 4; SEQ ID NO 34552; 650pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention
 XX
 XX Sequence 61 AA;
 SQ

Query Match 62.5%; Score 40; DB 4; Length 61;
 Best Local Similarity 72.7%; Pred. No. 5.3;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNHNGOK 11
 |||:||||
 Db 35 TSPKHNGOK 45

Search completed: September 15, 2004, 09:56:50
 Job time : 124 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2004, 09:54:51 ; Search time 32 seconds
(without alignments)
19.360 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNIHNGQKL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PGTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	59.4	82	4	US-09-543-681A-8211
2	38	59.4	467	4	US-09-134-001C-3235
3	37	57.8	255	4	US-09-899-039A-8876
4	37	57.8	358	4	US-09-252-991A-17017
5	37	57.8	838	4	US-09-235-451-2
6	37	57.8	838	4	US-09-132-316-3
7	37	57.8	838	4	US-09-667-422-9
8	36	56.2	191	4	US-09-252-991A-21437
9	36	56.2	247	4	US-09-543-681A-5590
10	36	56.2	270	4	US-09-489-039A-11843
11	36	56.2	630	1	US-08-487-890A-113
12	36	56.2	630	2	US-08-478-435-113
13	36	56.2	630	2	US-08-337-483-113
14	36	56.2	630	2	US-08-478-373-113
15	36	56.2	630	3	US-08-474-671-113
16	36	56.2	630	3	US-08-483-577A-113
17	36	56.2	630	3	US-08-897-438-113
18	36	56.2	630	4	US-08-637-654-113
19	36	56.2	630	4	US-08-649-518-113
20	35	54.7	147	4	US-09-328-352-7278
21	35	54.7	352	4	US-09-543-681A-5596
22	35	54.7	393	4	US-09-252-991A-17136
23	35	54.7	414	4	US-09-899-039A-11717
24	35	54.7	1088	4	US-09-920-804-2
25	34	53.1	120	6	5514582-36
26	34	53.1	131	4	US-09-621-976-4805
27	34	53.1	216	4	US-09-543-681A-7321

28	34	53.1	260	4	US-09-134-001C-5363	Sequence 5363, Ap
29	34	53.1	274	1	US-08-256-964A-19	Sequence 19, Appl
30	34	53.1	287	4	US-09-800-729-114	Sequence 114, App
31	34	53.1	287	4	US-09-800-729-153	Sequence 153, App
32	34	53.1	310	4	US-09-976-594-807	Sequence 807, App
33	34	53.1	504	4	US-09-252-991A-28279	Sequence 28279, A
34	34	53.1	613	4	US-09-800-729-92	Sequence 82, Appl
35	34	53.1	613	4	US-09-800-729-98	Sequence 98, Appl
36	33.5	52.3	233	4	US-09-540-236-4657	Sequence 2657, Ap
37	33	51.6	158	3	US-09-215-221-22	Sequence 22, Appl
38	33	51.6	221	3	US-09-247-373B-54	Sequence 54, Appl
39	33	51.6	278	4	US-09-145-828A-11	Sequence 11, Appl
40	33	51.6	278	4	US-09-903-456-18	Sequence 18, Appl
41	33	51.6	362	4	US-09-296-840A-2	Sequence 2, Appli
42	33	51.6	417	4	US-09-489-039A-9729	Sequence 9729, Ap
43	33	51.6	476	4	US-09-489-039A-7607	Sequence 7607, Ap
44	33	51.6	715	3	US-09-215-221-25	Sequence 25, Appl
45	33	51.6	857	4	US-09-425-335-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-543-681A-8211
; Sequence 8211, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8211
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8211

Query Match

Best Local Similarity 59.4%; Score 38; DB 4; Length 82;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPLNIHNGQKL 12

Db 29 TSPIPTNGSKL 40

RESULT 2

US-09-134-001C-3235
; Sequence 3235, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3235
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3235

```
Query Match      59.4%; Score 38; DB 4; Length 467;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 NIHQKQL 12
Db 456 NIHQKQL 463

RESULT 3
US-09-489-039A-8876
; Sequence 8876, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8876
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8876

Query Match      57.8%; Score 37; DB 4; Length 255;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SPLNIHQKQL 12
Db 71 TPNHIQAQNL 81

RESULT 4
US-09-252-991A-17017
; Sequence 17017, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17017
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17017

Query Match      57.8%; Score 37; DB 4; Length 358;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SPLNIHQKQL 12
Db 193 SPLRIHQGQL 203

RESULT 5
US-09-235-451-2
; Sequence 2, Application US/09235451
```

```
; GENERAL INFORMATION:
; APPLICANT: Julius, David J.
; APPLICANT: Caterina, Michael J.
; APPLICANT: Brake, Anthony J.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
; FILE REFERENCE: 9076/084CIP
; CURRENT APPLICATION NUMBER: US/09/235,451
; CURRENT FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: R. rattus
US-09-235-451-2

Query Match      57.8%; Score 37; DB 4; Length 838;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LNIHQNGQ 10
Db 163 LNIHQNGQ 169

RESULT 6
US-09-132-316-3
; Sequence 3, Application US/09132316B
; Patent No. 6444440
; GENERAL INFORMATION:
; APPLICANT: Young, Paul E.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Vanilloid Receptor-2
; FILE REFERENCE: 1488.1110000
; CURRENT APPLICATION NUMBER: US/09/132,316B
; CURRENT FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: US 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-132-316-3

Query Match      57.8%; Score 37; DB 4; Length 838;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LNIHQNGQ 10
Db 163 LNIHQNGQ 169

RESULT 7
US-09-667-422-9
; Sequence 9, Application US/09667422
; Patent No. 6482611
; GENERAL INFORMATION:
; APPLICANT: Cortright, Daniel
; APPLICANT: Krause, James
; TITLE OF INVENTION: Human Capsaicin Receptor and Uses Thereof
; FILE REFERENCE: HCR
; CURRENT APPLICATION NUMBER: US/09/667,422
```

```
; CURRENT FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Rattus sp.
; PUBLICATION INFORMATION:
; AUTHORS: Caterina, Michael J.
; AUTHORS: Schumacher, Mark A.
; AUTHORS: Tominaga, Makoto
; AUTHORS: Rosen, Tobias A.
; TITLE: The capsaicin receptor: a heat-activated ion channel in
; TITLE: the pain pathway
; JOURNAL: Nature
; VOLUME: 389
; PAGES: 816-824
; DATE: 1997
; US-09-867-422-9

Query Match      57.8%; Score 37; DB 4; Length 838;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      4 LNINHGQ 10
      ||:||||
Db      163 LNLHGQ 169

RESULT 8
US-09-252-991A-21437
; Sequence 21437, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21437
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21437

Query Match      56.2%; Score 36; DB 4; Length 191;
Best Local Similarity 77.8%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      4 LNINHGQKL 12
      |||||:|
Db      145 LNIHGREFL 153

RESULT 9
US-09-543-681A-5590
; Sequence 5590, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128.706
; PRIOR FILING DATE: 1999-04-09

; CURRENT FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-5590

Query Match      56.2%; Score 36; DB 4; Length 247;
Best Local Similarity 63.6%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 SPLNIHGQKL 12
      |||||:|
Db      80 SPLDLHLGQVL 90

RESULT 10
US-09-489-039A-11843
; Sequence 11843, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117.747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11843
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-11843

Query Match      56.2%; Score 36; DB 4; Length 270;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 SPLNIHGQ 10
      |||||:|
Db      103 SPLHIHLGQ 111

RESULT 11
US-08-487-890A-113
; Sequence 113, Application US/08487890A
; Patent No. 5708149
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Fele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/487,890A
FILING DATE: 07-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-890A-113

Query Match 56.2%; Score 36; DB 1; Length 630;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGQK 11
|:|:|:|:|:
Db 119 TNPLEKHGQR 129

RESULT 12

US-08-478-435-113
Sequence 113, Application US/08478435
Patent No. 5922323

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,435
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
APPLICATION NUMBER: US 08/148,968

FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-435-113

Query Match 56.2%; Score 36; DB 2; Length 630;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGQK 11
|:|:|:|:|:
Db 119 TNPLEKHGQR 129

RESULT 13

US-08-337-483-113
Sequence 113, Application US/08337483
Patent No. 5922562

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-337-483-113

Query Match 56.2%; Score 36; DB 2; Length 630;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNHNQOK 11
Db 119 TNPLEKHGQR 129

RESULT 14

US-08-478-373-113
; Sequence 113, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-373-113

Query Match 56.2%; Score 36; DB 2; Length 630;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNHNQOK 11
Db 119 TNPLEKHGQR 129

RESULT 15

US-08-474-671-113
; Sequence 113, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-671-113

Query Match 56.2%; Score 36; DB 3; Length 630;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNHNQOK 11
Db 119 TNPLEKHGQR 129

Search completed: September 15, 2004, 10:00:40
Job time: 33 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2004, 09:59:27 ; Search time 127 Seconds
(without alignments)
30.301 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPLNIHQKQL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 segs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	12	9 US-09-899-376-1	Sequence 1, Appli
2	64	100.0	19	9 US-09-899-376-3	Sequence 3, Appli
3	64	100.0	19	9 US-09-899-376-4	Sequence 4, Appli
4	42	65.6	1325	9 US-09-741-669-304	Sequence 304, App
5	40	62.5	61	9 US-09-864-761-40714	Sequence 40714, A
6	40	62.5	214	12 US-10-425-114-71139	Sequence 71139, A
7	40	62.5	408	15 US-10-369-493-15751	Sequence 15751, A
8	40	62.5	408	15 US-10-369-493-16135	Sequence 16135, A
9	40	62.5	442	15 US-10-369-493-15383	Sequence 15383, A
10	40	62.5	631	16 US-10-437-963-125370	Sequence 125370, A
11	39	60.9	243	12 US-10-282-122A-77766	Sequence 77766, A
12	39	60.9	610	15 US-10-094-749-2604	Sequence 2604, App
13	39	60.9	610	15 US-10-108-260A-3772	Sequence 3772, App
14	39	60.9	738	16 US-10-408-765A-740	Sequence 740, App
15	39	60.9	775	16 US-10-437-963-151193	Sequence 151193, A

Sequence 236481, Sequence 26, Appli
Sequence 219, App
Sequence 70605, A
Sequence 4, Appli
Sequence 2188, App
Sequence 68160, A
Sequence 5066, App
Sequence 43542, A
Sequence 59671, A
Sequence 4077, App
Sequence 17, Appli
Sequence 72, Appli
Sequence 3, Appli
Sequence 38, Appli
Sequence 40, Appli
Sequence 15500, A
Sequence 7756, App
Sequence 11660, A
Sequence 279801, A
Sequence 222, App
Sequence 7757, App
Sequence 56063, A
Sequence 59653, A
Sequence 69000, A
Sequence 76312, A
Sequence 10296, A
Sequence 56679, A
Sequence 73177, A

US-10-424-599-236481
US-10-042-417-36
US-10-043-487-219
US-10-282-122A-70605
US-10-729-807-4
US-10-369-493-2188
US-10-282-122A-68160
US-09-815-242-5066
US-10-282-122A-43542
US-10-369-493-4077
US-10-027-828-17
US-10-342-844-72
US-09-978-303-2
US-10-137-316-3
US-10-342-844-38
US-10-342-844-40
US-09-815-242-11500
US-10-335-977-7756
US-09-815-242-11660
US-10-424-599-279801
US-09-882-227-222
US-10-335-977-7757
US-10-282-122A-56063
US-10-282-122A-59653
US-10-282-122A-69000
US-10-282-122A-76312
US-09-815-242-10296
US-10-282-122A-56679
US-10-282-122A-73177

ALIGNMENTS

RESULT 1
US-09-899-376-1
; Sequence 1, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; TITLE OF INVENTION: INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
; FILE REFERENCE: UTSC:645US
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; TYPE: PRT
; LENGTH: 12
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-1

Query Match 100.0%; Score 64; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 6.2e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNIHQKQL 12
|||
Db 1 TSPLNIHQKQL 12

RESULT 2
US-09-899-376-3
; Sequence 3, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.

; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: US/09/899,376
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-3

Query Match 100.0%; Score 64; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGQKL 12
Db 4 TSPLNTHNGQKL 15

RESULT 3
US-09-899-376-4
; Sequence 4, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: US/09/899,376
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-4

Query Match 100.0%; Score 64; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGQKL 12
Db 8 TSPLNTHNGQKL 19

RESULT 4
US-09-741-669-304
; Sequence 304, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 304
; LENGTH: 1325
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-304

Query Match 65.6%; Score 42; DB 9; Length 1325;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LNTHNGQKL 12
Db 991 LNTHNGQKL 999

RESULT 5
US-09-864-761-40714
; Sequence 40714, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Rank, Sharron G.
; APPLICANT: Penn, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40714
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: MAP TO U82670.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
OTHER INFORMATION: SWISSPROT HIT: P94598, EVALUATE 1.50e+00
US-09-864-761-40714

Query Match 62.5%; Score 40; DB 9; Length 61;
Best Local Similarity 72.7%; Pred. No. 9.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPKIHNGQK 11
|||:|||||
Db 35 TSPKIHNGQK 45

RESULT 6
US-10-425-114-71139
Sequence 71139, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingsong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 71139
LENGTH: 214
TYPE: PRT
ORGANISM: Zea mays subsp. mexicana
FEATURE:
OTHER INFORMATION: Clone ID: UC-2MROTEOSINTE014G02_FLI.pep
US-10-425-114-71139

Query Match 62.5%; Score 40; DB 12; Length 214;
Best Local Similarity 70.0%; Pred. No. 39;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 PLNTHNGQKL 12
|||:|||||
Db 60 PLDHLNGLKV 69

RESULT 7
US-10-369-493-15751
Sequence 15751, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 15751
LENGTH: 408
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-10-369-493-15751

Query Match 62.5%; Score 40; DB 15; Length 408;
Best Local Similarity 58.3%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSPLNIHNGQKL 12
|||:|||||
Db 135 TPPLRVHGGQPL 146

RESULT 8
US-10-369-493-16135
Sequence 16135, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 16135
LENGTH: 408
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-10-369-493-16135

Query Match 62.5%; Score 40; DB 15; Length 408;
Best Local Similarity 58.3%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSPLNIHNGQKL 12
|||:|||||
Db 135 TPPLRVHGGQPL 146

RESULT 9
US-10-369-493-15383
Sequence 15383, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 15383
LENGTH: 442
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-10-369-493-15383

Query Match 62.5%; Score 40; DB 15; Length 442;

```

Best Local Similarity 58.3%; Pred. No. 86;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSPLNHNHNGKL 12
    |||:||||
Db 158 TPPLRVHGGQPL 169

RESULT 10
US-10-437-963-125370
; Sequence 125370, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 125370
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2801C.1.1.dep
US-10-437-963-125370

Query Match 62.5%; Score 40; DB 16; Length 631;
Best Local Similarity 72.7%; Pred. No. 1.3e-02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNHNHNGOK 11
    |||:||||
Db 470 TSPLAHQNGKK 480

RESULT 11
US-10-282-122A-77766
; Sequence 77766, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06

```

```

; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 77766
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Versinia pestis
US-10-282-122A-77766

Query Match 60.9%; Score 39; DB 12; Length 243;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNHNHNGQ 10
    |||:||||
Db 76 SPLNHLHGQ 84

RESULT 12
US-10-094-749-2604
; Sequence 2604, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2604
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2604

Query Match 60.9%; Score 39; DB 15; Length 610;
Best Local Similarity 77.8%; Pred. No. 1.9e-02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 4 LNIHNGOKL 12
| | | | | | | |
Db 377 LRIHNGEKL 385

RESULT 13

US-10-108-260A-3772
; Sequence 3772, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3772
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3772

Query Match 60.9%; Score 39; DB 15; Length 610;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGOKL 12
| | | | | | | |
Db 377 LRIHNGEKL 385

RESULT 14

US-10-408-765A-740
; Sequence 740, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 740
; LENGTH: 738
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-740

Query Match 60.9%; Score 39; DB 16; Length 738;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGOKL 12
| | | | | | | |
Db 505 LRIHNGEKL 513

RESULT 15

US-10-437-963-151193
; Sequence 151193, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 151193
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_51359C.1.pép
US-10-437-963-151193

Query Match 60.9%; Score 39; DB 16; Length 775;
Best Local Similarity 70.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PLNIHNGOKL 12
| | | | | | | |
Db 752 PLRIHNGKTL 761

Search completed: September 15, 2004, 10:11:04
Job time : 128 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2004, 09:56:56 ; Search time 413 Seconds
(without alignments)
28.360 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TGPLNTHNQKL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
------------	-------------	--------------	-------	-------------

1	64	100.0	12	23	US-09-899-376-1	Sequence 1, Appli
2	64	100.0	12	33	US-60-492-508-32	Sequence 32, Appli
3	64	100.0	19	23	US-09-899-376-3	Sequence 3, Appli
4	64	100.0	19	23	US-09-899-376-4	Sequence 4, Appli
5	42	65.6	84	33	US-60-160-203-3860	Sequence 3860, Ap
6	42	65.6	84	33	US-60-160-209-2774	Sequence 2774, Ap
7	42	65.6	98	33	US-60-169-840-5692	Sequence 5692, Ap
8	42	65.6	98	33	US-60-169-868-4082	Sequence 4082, Ap
9	42	65.6	258	1	PCT-US02-36123-4264	Sequence 4264, Ap
10	42	65.6	300	22	US-09-791-537-84058	Sequence 84058, A
11	42	65.6	470	1	PCT-US01-14827-11967	Sequence 11967, A
12	42	65.6	804	1	PCT-US01-08631-59159	Sequence 59159, A
13	42	65.6	804	1	PCT-US01-08631-59382	Sequence 59382, A
14	42	65.6	807	1	PCT-US01-08631-59024	Sequence 59024, A
15	42	65.6	1023	1	PCT-US01-08631-59027	Sequence 59027, A
16	42	65.6	1302	20	US-09-614-150-21549	Sequence 21549, A
17	42	65.6	1302	33	US-09-614-150A-21549	Sequence 21549, A
18	42	65.6	1302	33	US-60-167-245-768	Sequence 768, App
19	42	65.6	1302	33	US-60-173-464-17708	Sequence 17708, A
20	42	65.6	1302	33	US-60-191-637-21614	Sequence 21614, A
21	42	65.6	1302	33	US-60-191-681-17042	Sequence 17042, A
22	42	65.6	1325	21	US-09-741-669-304	Sequence 304, App
23	41	64.1	281	21	US-09-708-427-28206	Sequence 28206, A
24	41	64.1	326	21	US-09-708-427-28205	Sequence 28205, A
25	41	64.1	427	19	US-09-573-655A-2238	Sequence 2238, Ap
26	41	64.1	427	19	US-09-573-655B-2238	Sequence 2238, Ap
27	41	64.1	427	21	US-09-708-427-28204	Sequence 28204, A
28	40	62.5	61	1	PCT-US01-00663-35637	Sequence 35637, A
29	40	62.5	61	23	US-09-864-761-40714	Sequence 40714, A
30	40	62.5	61	27	US-10-182-993-34552	Sequence 34552, A
31	40	62.5	61	27	US-10-182-993-34552	Sequence 34552, A
32	40	62.5	61	28	US-10-182-995-27186	Sequence 27186, A
33	40	62.5	61	28	US-10-203-134-35562	Sequence 35562, A
34	40	62.5	61	28	US-10-203-135-34569	Sequence 34569, A
35	40	62.5	61	28	US-10-203-136-35666	Sequence 35666, A
36	40	62.5	61	28	US-10-203-137-35637	Sequence 35637, A
37	40	62.5	61	28	US-10-203-139-34211	Sequence 34211, A
38	40	62.5	101	22	US-09-758-449-865	Sequence 865, App
39	40	62.5	101	28	US-10-212-778-865	Sequence 865, App
40	40	62.5	214	30	US-10-219-999-63024	Sequence 63024, A
41	40	62.5	214	30	US-10-425-114-71139	Sequence 71139, A
42	40	62.5	214	33	US-60-324-109-27105	Sequence 27105, A
43	40	62.5	214	33	US-10-369-493-15751	Sequence 15751, A
44	40	62.5	408	29	US-10-369-493-16135	Sequence 16135, A
45	40	62.5	408	33	US-60-360-039-15751	Sequence 15751, A

ALIGNMENTS

RESULT 1
US-09-899-376-1
; Sequence 1, Application US/09899376
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
; FILE REFERENCE: UTSC:645US
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-1

Query Match 100.0%; Score 64; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.00017; Mismatches 0; Indels 0; Gaps 0;
Matches 12; Conservative 0;

Qy 1 TSPLNHNHNGQKL 12
Db 1 TSPLNHNHNGQKL 12

RESULT 2

US-60-492-508-32
; Sequence 32, Application US/60492508
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; TITLE OF INVENTION: TARGETED CARRIER FUSIONS FOR DELIVERY OF
; FILE REFERENCE: WARP:01IUS1
; CURRENT APPLICATION NUMBER: US/60/492,508
; CURRENT FILING DATE: 2003-08-05
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 32
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-60-492-508-32

Query Match 100.0%; Score 64; DB 33; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNHNHNGQKL 12
Db 1 TSPLNHNHNGQKL 12

RESULT 3

US-09-899-376-3
; Sequence 3, Application US/09899376
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: UTSC:645US
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-3

Query Match 100.0%; Score 64; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNHNHNGQKL 12
Db 4 TSPLNHNHNGQKL 15

RESULT 4

US-09-899-376-4
; Sequence 4, Application US/09899376
; GENERAL INFORMATION:

; APPLICANT: HONG, FRANK D.
; APPLICANT: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: UTSC:645US
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-4

Query Match 100.0%; Score 64; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNHNHNGQKL 12
Db 8 TSPLNHNHNGQKL 19

RESULT 5

US-60-160-203-3860
; Sequence 3860, Application US/60160203
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIAN
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; FILE REFERENCE: CLC000116
; CURRENT APPLICATION NUMBER: US/60/160,203
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 6374
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3860
; LENGTH: 84
; TYPE: PRT
; ORGANISM: HUMAN
US-60-160-203-3860

Query Match 65.6%; Score 42; DB 33; Length 84;
Best Local Similarity 72.7%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TSPLNHNHNGQKL 11
Db 64 TSPLNHNHNGQKL 74

RESULT 6

US-60-160-209-2774
; Sequence 2774, Application US/60160209
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIAN
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; FILE REFERENCE: CLC000113
; CURRENT APPLICATION NUMBER: US/60/160,209
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 4646
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2774
; LENGTH: 84
; TYPE: PRT
; ORGANISM: HUMAN
US-60-160-209-2774

```
Query Match      65.6%; Score 42; DB 33; Length 84;
Best Local Similarity 72.7%; Pred. No. 20;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNIHNGQK 11
      ||||| :
Db      64 TSPLTIHNVK 74

RESULT 7
US-60-169-840-5692
; Sequence 5692, Application US/60169840
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00164
; CURRENT APPLICATION NUMBER: US/60/169,840
; CURRENT FILING DATE: 1999-12-09
; NUMBER OF SEQ ID NOS: 9628
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5692
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Human
US-60-169-840-5692

Query Match      65.6%; Score 42; DB 33; Length 98;
Best Local Similarity 72.7%; Pred. No. 24;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNIHNGQK 11
      ||||| :
Db      76 TSPLTIHNVK 86

RESULT 8
US-60-169-868-4082
; Sequence 4082, Application US/60169868
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00161
; CURRENT APPLICATION NUMBER: US/60/169,868
; CURRENT FILING DATE: 1999-12-09
; NUMBER OF SEQ ID NOS: 6938
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4082
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Human
US-60-169-868-4082

Query Match      65.6%; Score 42; DB 33; Length 98;
Best Local Similarity 72.7%; Pred. No. 24;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNIHNGQK 11
      ||||| :
Db      69 TSPLTIHNVK 79

RESULT 9
PCT-US02-36123-4264
; Sequence 4264, Application PC/TUS0236123
; GENERAL INFORMATION:
; APPLICANT: American Cyanamid Company, and Fletcher, Leah D., McMichael, John C.,
; APPLICANT: Russell, David P., and Zagursky, Robert J.
; TITLE OF INVENTION: Alloiococcus otitidis Open Reading Frames (ORFs) Encoding Polypep
```

```
; TITLE OF INVENTION: Antigens, Immunogenic Compositions and Uses Thereof
; FILE REFERENCE: Application 1
; CURRENT APPLICATION NUMBER: PCT/US02/36123
; CURRENT FILING DATE: 2003-01-02
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4264
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
PCT-US02-36123-4264

Query Match      65.6%; Score 42; DB 1; Length 258;
Best Local Similarity 58.3%; Pred. No. 70;
Matches      7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNIHNGQKL 12
      ||||| :
Db      10 SSPLEIHGGRKM 21

RESULT 10
US-09-791-537-84058
; Sequence 84058, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 84058
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-09-791-537-84058

Query Match      65.6%; Score 42; DB 22; Length 300;
Best Local Similarity 58.3%; Pred. No. 83;
Matches      7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 TSPLNIHNGQKL 12
      ||||| :
Db      122 TAPLNVIGGTKL 133

RESULT 11
PCT-US01-14827-11967
; Sequence 11967, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 11967
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (260)..(450)
; OTHER INFORMATION: Transposase domain identified by Pfam, accession name
; OTHER INFORMATION: Transposase_15, E-value=7.4e-90, Pfam score of 311.9
PCT-US01-14827-11967
```

```

Query Match      65.6%; Score 42; DB 1; Length 470;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

QY      4 LNIHNGQKL 12
      |||||
Db      47 LNIHGGQKL 55

RESULT 12
PCT-US01-08631-59159
; Sequence 59159, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59159
; LENGTH: 804
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (643)..(671)
; OTHER INFORMATION: NADH dehydrogenases domain identified by Pfam, accession name
; OTHER INFORMATION: NADHdh, E-value=4.8e-07, Pfam score of 28.7
PCT-US01-08631-59159

Query Match      65.6%; Score 42; DB 1; Length 804;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

QY      4 LNIHNGQKL 12
      |||||
Db      325 LNIHGGQKL 333

RESULT 13
PCT-US01-08631-59382
; Sequence 59382, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59382
; LENGTH: 804
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (643)..(671)
; OTHER INFORMATION: NADH dehydrogenases domain identified by Pfam, accession name
; OTHER INFORMATION: NADHdh, E-value=4.8e-07, Pfam score of 28.7
PCT-US01-08631-59382

Query Match      65.6%; Score 42; DB 1; Length 804;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

```

```

Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

QY      4 LNIHNGQKL 12
      |||||
Db      325 LNIHGGQKL 333

RESULT 14
PCT-US01-08631-59024
; Sequence 59024, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59024
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (646)..(674)
; OTHER INFORMATION: NADH dehydrogenases domain identified by Pfam, accession name
; OTHER INFORMATION: NADHdh, E-value=4.8e-07, Pfam score of 28.7
PCT-US01-08631-59024

Query Match      65.8%; Score 42; DB 1; Length 807;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches      8; Conservative      0; Mismatches      1; Indels      0; Gaps      0;

QY      4 LNIHNGQKL 12
      |||||
Db      328 LNIHGGQKL 336

RESULT 15
PCT-US01-08631-59027
; Sequence 59027, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59027
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (179)..(202)
; OTHER INFORMATION: Dihydroxy-acid and 6-phosphogluconate dehydratases proteins
; OTHER INFORMATION: domain identified by eMATRIX, accession number BL00866F, p-value=
; OTHER INFORMATION: 2.22e-24, raw score of 13.15
; NAME/KEY: DOMAIN
; LOCATION: (689)..(802)
; OTHER INFORMATION: Response regulator receiver domain identified by Pfam,
; OTHER INFORMATION: accession name response_reg, E-value=1.3e-42, Pfam score of 155.0
PCT-US01-08631-59027

```


Query Match 65.6%; Score 42; DB 1; Length 1023;
Best Local Similarity 88.9%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LNIHNGQKL 12
|||
Db 439 LNIHGGQKL 447

Search completed: September 15, 2004, 10:07:40
Job time : 414 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

- OM protein - protein search, using sw model

Run on: September 15, 2004, 09:53:40 ; Search time 39 Seconds
(without alignments)
29.597 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNIHNGQKL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	68.8	1302	2 B41249	multidrug resistan
2	42	65.6	300	2 S41535	ribC protein - Shi
3	42	65.6	1325	2 A64905	ydcB protein - Esc
4	42	65.6	1343	2 B90893	hypothetical prote
5	42	65.6	1343	2 D85724	hypothetical prote
6	42	65.6	2531	2 T31070	notch homolog - se
7	41	64.1	427	2 T08980	hypothetical prote
8	40	62.5	331	2 T38546	hypothetical proli
9	39	60.9	243	2 AF0114	conserved hypothet
10	38	59.4	563	2 T36580	hypothetical prote
11	38	59.4	150	2 F86635	transporter yajA l
12	38	59.4	223	2 F31201	GLI-related finger
13	38	59.4	324	2 T24465	hypothetical prote
14	38	59.4	406	2 AD0833	probable membrane
15	38	59.4	441	2 T29991	hypothetical prote
16	38	59.4	683	2 T38254	serine/threonine-s
17	38	59.4	693	2 C84495	hypothetical prote
18	37	57.8	111	2 A35137	suppressor protein
19	37	57.8	111	2 G91129	probable proteinase
20	37	57.8	111	2 C85974	conserved hypothet
21	37	57.8	240	2 F83592	hypothetical prote
22	37	57.8	248	2 S23348	ribA protein - Sal
23	37	57.8	299	2 S15300	hypothetical prote
24	37	57.8	299	2 AE0767	dTDP-4-dehydrozham
25	37	57.8	476	1 FOLUBT	gag polyprotein -
26	37	57.8	587	2 S58319	hypothetical prote
27	37	57.8	838	2 T09054	capsaicin receptor
28	36	56.2	169	2 G83075	type 4 fibrial bl
29	36	56.2	194	2 H71939	hypothetical prote

30 36 56.2 217 2 B64651 hypothetical prote
31 36 56.2 243 2 AF0877 conserved hypothet
32 36 56.2 252 2 A65080 hypothetical prote
33 36 56.2 252 2 F91106 hypothetical prote
34 36 56.2 252 2 A85952 hypothetical prote
35 36 56.2 253 2 G84977 hypothetical prote
36 36 56.2 265 1 I40649 hypothetical prote
37 36 56.2 307 2 S75208 dhaj protein - syn
38 36 56.2 372 2 T25717 hypothetical prote
39 36 56.2 454 2 A97048 similar to phospho
40 36 56.2 605 2 H71303 hypothetical prote
41 36 56.2 630 2 S70909 transfeirin-bindin
42 36 56.2 817 2 T21336 beta-galactosidase
43 36 56.2 1005 2 T31333 hypothetical prote
44 36 56.2 1474 2 B85188 retroransposon li
45 35 54.7 15 2 A60929 dichloromethane de

ALIGNMENTS

RESULT 1
B41249
multidrug resistance protein homolog Mdr65 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 03-Apr-1992 #sequence_revision 12-Jun-1992 #text_change 02-Feb-2001
C:Accession: B41249
R:Wu, C.T.; Budding, M.; Griffin, M.S.; Croop, J.M.
Mol. Cell. Biol. 11, 3940-3948, 1991
A:Title: Isolation and characterization of Drosophila multidrug resistance gene homolog;
A:Reference number: A41249; MUID:91304385; PMID:2072901
A:Accession: B41249
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1302 <WUA>
A:Cross-references: GB:M59077; NID:g157874; PIDN:AAA28680.1; PID:g157875
C:Genetics:
A:Gene: FlyBase:Mdr65
A:Cross-references: FlyBase:FBgn0004513
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
F:423-617/Domain: ATP-binding cassette homology <ABC1>
F:440-447/Region: nucleotide-binding motif A (P-loop)
F:1077-1274/Domain: ATP-binding cassette homology <ABC2>
F:1094-1101/Region: nucleotide-binding motif A (P-loop)

Query Match 68.8%; Score 44; DB 2; Length 1302;
Best Local Similarity 72.7%; Pred. No. 7.1;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNIHNGQK 11
DB 673 TSPLNIHNGQK 693

RESULT 2
S41535
rfbC protein - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999
C:Accession: S41535
R:Macpherson, D.F.; Manning, P.A.; Morona, R.
Mol. Microbiol. 11, 281-292, 1994
A:Title: Characterization of the dTDP-rhamnose biosynthetic genes encoded in the rfb lo
A:Reference number: S41533; MUID:94224146; PMID:8170390
A:Accession: S41535
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <MAC>
A:Cross-references: EMBL:X71970; NID:G506557; PIDN:CAA50768.1; PID:G454899
C:Superfamily: dTDP-dihydrostreptose synthase

Query Match 65.6%; Score 42; DB 2; Length 300;

Best Local Similarity 58.3%; Pred. No. 3.2;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNTHNGQKL 12
|:|:|:|:|
Db 122 TAPLNVGQTKL 133

RESULT 3
A64905
ydek protein - Escherichia coli (strain K-12)
N;Alternate names: protein T
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: A64905; 152440; S3415 #text_change 01-Mar-2002
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64905
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1325 <BLAT>
A;Cross-references: GB:AE000248; GB:U00096; NID:G1787783; PIDN:AAC74583.1; PID:G1787788;
A;Experimental source: strain K-12, substrain MG1655
R;Cartwright, P.; Timms, M.; Lithgow, T.; Hoj, P.; Hoogenraad, N.
Biochim. Biophys. Acta 1153, 345-347, 1993
A;Title: An Escherichia coli gene showing a potential ancestral relationship to the gene
A;Reference number: 152440; MUID:94100243; PMID:8274505
A;Accession: 152440
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 689-883, 'K', 885-1316, 'S', 1318-1325 <RES>
A;Cross-references: EMBL:X73295; NID:G312392; PIDN:CAAS1730.1; PID:G312393
A;Note: the difference in length is due to a frameshift error at pos 653
C;Genetics:
A;Gene: ydek
C;Function:
A;Description: probably involved in protein translocation apparatus
C;Keywords: nucleotide binding; P-loop
F;712-719/Region: nucleotide-binding motif A (P-loop)

Query Match 65.6%; Score 42; DB 2; Length 1325;
Best Local Similarity 88.9%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNTHNGQKL 12
|:|:|:|:|
Db 991 LNTHNGQKL 999

RESULT 4
E90893
hypothetical protein ECs2117 [imported] - Escherichia coli (strain O157:H7, substrain R1
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: E90893
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90893
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1343 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA35540.1; PID:G13361583; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain R1MD 0509952
C;Genetics:
A;Gene: ECs2117

Query Match 65.6%; Score 42; DB 2; Length 1343;

Best Local Similarity 88.9%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNTHNGQKL 12
|:|:|:|:|
Db 991 LNTHNGQKL 999

RESULT 5
D85724
hypothetical protein ydek [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85724
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1343 <STO>
A;Cross-references: GB:AE005174; NID:G12515159; PIDN:AAGS6256.1; GSPDB:GN00145; UMG:221
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ydek

Query Match 65.6%; Score 42; DB 2; Length 1343;
Best Local Similarity 88.9%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNTHNGQKL 12
|:|:|:|:|
Db 991 LNTHNGQKL 999

RESULT 6
T31070
notch homolog - sea urchin (Lytechinus variegatus)
C;Species: Lytechinus variegatus (variegated urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C;Accession: T31070
R;Sherwood, D.R.; McClay, D.R.
Development 124, 3363-3374, 1997
A;Title: Identification and localization of a sea urchin Notch homologue: insights into
A;Reference number: Z20966; MUID:97454256; PMID:9310331
A;Accession: T31070
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2531 <SHE>
A;Cross-references: EMBL:AF000634; NID:G2570350; PID:G2570351; PIDN:AAB82088.1
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 65.6%; Score 42; DB 2; Length 2531;
Best Local Similarity 60.0%; Pred. No. 36;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNTHNGQ 10
|:|:|:|:|
Db 2097 TSPMDWHNGE 2106

RESULT 7
T08980
hypothetical protein F6G3.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cross)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08980
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16520
A;Accession: T08980

Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPNINHNGK 11
| | | | |
Db 112 TQHLRIHNGEK 122

RESULT 13

T24465

hypothetical protein T04F8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T24465

R;Jennard, N.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19895

A:Accession: T24465

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-324 <WIL>

A:Cross-references: EMBL:Z66565; PIDN:CAA91477.1; GSPDB:GN000028; CESP:T04F8.1

A:Experimental source: clone T04F8

C:Genetics:

A:Gene: CESP:T04F8.1

A:Map position: X

A:Introns: 29/3; 58/2; 172/3; 201/2; 244/1

C:Superfamily: Saccharomyces probabie membrane protein YOR271c

Query Match

59.4%; Score 38; DB 2; Length 324;

Best Local Similarity 54.5%; Pred. No. 20;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSPNINHNGK 11

| | | | |

Db 35 TNPINLFHGEK 45

RESULT 14

AD0833

Probable membrane protein STY2859 [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AD0833

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD0833

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-406 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05851.1; PID:gi6503826; GSPDB:GN00176

C:Genetics:

A:Gene: STY2859

Query Match

59.4%; Score 38; DB 2; Length 406;

Best Local Similarity 50.0%; Pred. No. 26;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 PLNINHNGK 12

| | | | |

Db 356 PPELHNGQRI 365

RESULT 15

T29991

hypothetical protein C43H6.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29991
R;Le, T.T.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid C43H6.

A:Reference number: Z2017

A:Accession: T29991

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-641 <LET>

A:Cross-references: EMBL:U51999; PIDN:AAA96085.1; GSPDB:GN000028; CESP:C43H6.4

A:Experimental source: strain Bristol N2; clone C43H6

C:Genetics:

A:Gene: CESP:C43H6.4

A:Map position: X

A:Introns: 12/1; 63/2; 89/3; 143/1; 187/3; 268/3; 364/3; 403/3; 442/2; 487/2; 517/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C43H6.4

Query Match

59.4%; Score 38; DB 2; Length 641;

Best Local Similarity 72.7%; Pred. No. 44;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPLNINHNGK 12

| | | | |

Db 510 SPLNIFKGYKL 520

Search completed: September 15, 2004, 10:00:03

Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2004, 09:46:10 ; Search time 23 Seconds
(without alignments)
27.167 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNTHNQKL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	65.6	1302	MDR5_DROME	Q00748 drosophila
2	42	65.6	1325	YDEK_ECOLI	P32051 escherichia
3	40	62.5	438	AROA_XANCP	Q8p955 xanthomonas
4	39	60.9	434	TOLB_CHLTE	C8keq0 chlorobium
5	39	60.9	939	HEX_ADEP3	Q9ytr8 porcine ade
6	38	59.4	321	Z177_HUMAN	Q13360 homo sapien
7	38	59.4	376	GLI4_HUMAN	P10075 homo sapien
8	38	59.4	471	FBX3_HUMAN	P10075 homo sapien
9	38	59.4	683	FLOI_SCHPO	P50528 schizosacch
10	37	57.8	111	SOHA_ECOLI	P15373 escherichia
11	37	57.8	299	RFB_D_SALTY	P26392 salmonella
12	37	57.8	476	GAG_BIV06	P19558 bovine immu
13	37	57.8	476	GAG_BIV27	P19558 bovine immu
14	36	56.2	194	MINC_HELPJ	Q25693 helicobacte
15	36	56.2	195	MINC_HELPY	Q25693 helicobacte
16	36	56.2	243	YGGJ_ECOLI	P37912 escherichia
17	36	56.2	253	Y410_BUCAI	P57488 buchneza ap
18	36	56.2	307	DNAJ_SYNY3	P73037 synchocyst
19	36	56.2	429	BIOA_BUCBP	Q89ax4 buchneza ap
20	36	56.2	430	SUN_COXBU	P45679 coxiella bu
21	36	56.2	605	Y598_TREPA	Q83607 treponema p
22	36	56.2	1005	BGAL_ACPPL	P70753 actinobacil
23	35	54.7	241	UNG_AGRYS	Q8umcd agrobacteri
24	35	54.7	288	DCMA_METSP	P21161 methylobact
25	35	54.7	436	GI01_HUMAN	Q8taif homo sapien
26	35	54.7	451	HN4A_XENLA	Q91766 xenopus lae
27	35	54.7	491	G6PD_BUCAI	P57405 buchneza ap
28	35	54.7	551	SYE_ARCFU	Q29979 archaeoglob
29	35	54.7	928	YDGI_SCHPO	Q10431 schizosacch
30	35	54.7	942	DP01_CHLAU	Q08307 chloroflexu
31	35	54.7	1826	K13E_HUMAN	Q9nqt8 homo sapien
32	35	54.7	5938	MAC4_HUMAN	Q96pk2 homo sapien
33	34	53.1	93	YGGJ_ERWCH	P37995 erwiania chr

34 34 53.1 205 1 LEXA_PROBE
35 34 53.1 223 1 GIX1_TOBAC
36 34 53.1 285 1 YIF0_YEAST
37 34 53.1 287 1 KDSA_CAUCR
38 34 53.1 299 1 RBD1_ECOLI
39 34 53.1 299 1 RBD1_SHIFL
40 34 53.1 301 1 RBD2_ECOLI
41 34 53.1 454 1 AROA_XYLFA
42 34 53.1 454 1 AROA_XYLFT
43 34 53.1 577 1 G6PC_SOUTU
44 34 53.1 672 1 OSM3_CABEL
45 34 53.1 721 1 BBS2_MOUSE

ALIGNMENTS

RESULT 1
MDR5_DROME
ID MDR5_DROME STANDARD; PRT; 1302 AA.
AC Q00748; Q9VRW3;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Multidrug resistance protein homolog 65 (P-glycoprotein 65).
GN MDR65 OR CG10181.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Head;
RX MEDLINE=91304385; PubMed=2072901;
RT Wu C.-I., Budding M., Griffin M.S., Croop J.M.;
RT "Isolation and characterization of Drosophila multidrug resistance
RT gene homologs.";
RL Mol. Cell. Biol. 11:3940-3948(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WS1127, and WS1125;
RX MEDLINE=20467314; PubMed=11012721;
RA Begun D.J., Whitley P.;
RT "Genetics of alpha-amanitin resistance in a natural population of
RT Drosophila melanogaster.";
RL Heredity 85:184-190(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Makos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabros B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Feiraz C., Ferreira S., Fleischmann W.,
RA Fostier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

Q07267 providencia
Q03662 nicotiana t
P40186 saccharomyc
Q9a8cs c 2-dehydro
P37760 escherichia
P37778 shigella fl
Q46769 escherichia
Q9b21 xyella fas
Q87bu2 xyella fas
P43839 solanum tub
P46873 caenorhabdi
Q9cwf6 mus musculu

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M59077; AAA28680.1; -;
 DR EMBL; AF251287; AAF69147.1; -;
 DR EMBL; AF251286; AAF69146.1; -;
 DR EMBL; AB003563; AAF50669.1; -;
 DR F1R; B41249; B41249.
 DR Flybase; FBgn0004513; Mdr65.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC TM transporter.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00664; ABC membrane; 2.
 DR Pfam; PF00005; ABC trans; 2.
 DR ProDom; PD000006; ABC transporter; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00229; ABC TM1F; 2.
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 2.
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 2.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 49 69 1 (POTENTIAL).
 FT DOMAIN 70 118 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 119 147 2 (POTENTIAL).
 FT DOMAIN 148 194 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 195 215 3 (POTENTIAL).
 FT DOMAIN 216 223 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 224 242 4 (POTENTIAL).
 FT DOMAIN 243 302 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 303 323 5 (POTENTIAL).
 FT DOMAIN 324 341 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 342 362 6 (POTENTIAL).
 FT DOMAIN 363 731 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 732 753 7 (POTENTIAL).
 FT DOMAIN 754 776 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 777 798 8 (POTENTIAL).
 FT DOMAIN 799 852 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 853 873 9 (POTENTIAL).
 FT DOMAIN 874 874 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 875 894 10 (POTENTIAL).
 FT DOMAIN 895 956 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 957 977 11 (POTENTIAL).
 FT DOMAIN 978 993 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 994 1014 12 (POTENTIAL).
 FT DOMAIN 1015 1302 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 440 447 ATP (POTENTIAL).
 FT NP_BIND 1094 1101 ATP (POTENTIAL).
 FT REPEAT 1 673 1.

FT REPEAT 674 1302 2.
 FT CARBOHYD 103 103 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 369 369 T -> S (IN REF. 1).
 FT CONFLICT 678 678 F -> L (IN REF. 1).
 SQ SEQUENCE 1302 AA; 143784 MW; 39A7BCBAFA31924A CRC64;
 Query Match 65.6%; Score 42; DB 1; Length 1302;
 Best Local Similarity 72.7%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TSPINHGCK 11
 DB 673 TSPINFGCK 683
 RESULT 2
 YDEK_ECOLI
 ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
 AC P32051; P76140; P77168;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical lipoprotein ydek precursor (ORF1).
 GN YDEK OR ORF1 OR B1510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE=94100243; PubMed=8274505;
 RA Cartwright P.J., Timms M.W., Lithgow T., Hoej P.B., Hoogenraad N.J.;
 RT "An *Escherichia coli* gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";
 RL Biochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (POTENTIAL).
 CC -1- SIMILARITY: TO E.COLI YFAL.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 ISP42 AND MOM38.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
 frameshift in position 653.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC


```

CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; A000248; AAC74583.1; -.
DR EMBL; D90793; BAA15190.1; ALT INIT.
DR EMBL; D90794; BAA15197.1; ALT INIT.
DR EMBL; X73295; CAA51730.1; ALT_FRAME.
DR PR; A64905; A64905.
DR EcGene; EGI1780; ydek.
DR InterPro; IPR000437; Prok lipoprot.S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal;
Complete proteome; Palmitate.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1325 HYPOTHETICAL LIPOPROTEIN YDEK.
FT LIPID 19 19 N-palmitoyl cysteine (Potential).
FT LIPID 19 19 S-diacylglycerol cysteine (Potential).
FT CONFLICT 884 884 N -> K (IN REF. 3).
FT CONFLICT 1317 1317 M -> S (IN REF. 3).
SQ SEQUENCE 1325 AA; 136514 MW; 26A3A066FA19AD7D CRC64;

Query Match 65.8%; Score 42; DB 1; Length 1325;
Best Local Similarity 88.9%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNTHGQKL 12
||| |||
Db 991 LNTHGQKL 999

RESULT 3
AROA_XANCP
ID AROA_XANCP STANDARD; PRT; 438 AA.
AC Q8PA95;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 3-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE encolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR XCC1591.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; Pubmed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergro F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Mencia C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Trufi D., Tsai S.M., White F.F.,
RA Scubal J.C., Kitajima J.F.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-((1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; A012260; AAM40886.1; -.
DR HAVAP; MF 00210; -.
DR InterPro; IPR001986; EPSP synth.
DR Pfam; PF00275; EPSP synthase; 1.
DR ProDom; PD001867; EPSP synthase; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 438 AA; 45100 MW; 6BF1EDCB80387AFA CRC64;

Query Match 62.5%; Score 40; DB 1; Length 438;
Best Local Similarity 58.3%; Pred. No. 9.2;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSPLNHNGQKL 12
||| |||
Db 150 TSPLEVHGQAL 161

RESULT 4
TOLB_CHLTE
ID TOLB_CHLTE STANDARD; PRT; 434 AA.
AC Q8KE00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE TOLB protein precursor.
GN TOLB OR CT0636.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; Pubmed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Hickey E.K., Brenner M., Shear T.P., Parksey D.,
RA Holt I., Umayam L.A., Mason T., Hansen C.L., Craven M.B., Radune D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Gruber T.M., Ketchum K.A.,
RA Vamathevan J., Khouri H., White O., Bryant D.A., Fraser C.M.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium."
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC -1- FUNCTION: Involved in the tonB-independent uptake of proteins (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: Belongs to the tolB family.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; A012837; AAM71875.1; -.
DR TIGR; CT0636; -.
DR HAVAP; MF 00671; -.
KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.

```

```

FT SIGNAL 1 27 Potential.
FT CHAIN 28 434 TolB protein.
SQ SEQUENCE 434 AA; 47275 MW; F1A347B89C7A0F99 CRC64;

Query Match 50.9%; Score 39; DB 1; Length 434;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNIHQG 10
Db 84 APLNIHQG 92

RESULT 5
HEX_ADEP3 STANDARD; PRT; 939 AA.
AC Q9YTR8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hexon protein (late protein 2).
GN P11.
OS Porcine adenovirus type 3 (PAV-3).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=35265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6618;
RX MEDLINE=99058191; PubMed=9837805;
RA Reddy P.S., Idamakanti N., Song J.-Y., Lee J.B., Hyun B.H., Park J.H.,
RA Cha S.H., Bae Y.T., Tikoo S.K., Babink L.A.;
RT "Nucleotide sequence and transcription map of porcine adenovirus type
RT 3.";
RL Virology 251:414-426(1998).
CC -!- FUNCTION: This protein is one of the structural proteins in the
CC viral coat and is synthesized during late infection.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; AF083132; AAC99441.1; -.
DR HSSP; P03277; IDHX.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR Pfam; PF03678; Adeno_hexon_C; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
KW Coat protein; Hexon protein; Late protein.
SQ SEQUENCE 939 AA; 106087 MW; 3B3B98EAC7C794EE CRC64;

Query Match 60.9%; Score 39; DB 1; Length 939;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SPLNIHQG 10
Db 223 SPTNIHQG 231

RESULT 6
Z177_HUMAN STANDARD; PRT; 321 AA.
AC Q13360;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 177.
GN ZNF177.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96299641; PubMed=8661005;
RA Baban S., Freeman J.D., Mager D.L.;
RT "Transcripts from a novel human KRAB zinc finger gene contain spliced
RT Alu and endogenous retroviral segments.";
RL Genomics 33:463-472(1996).
CC -!- FUNCTION: May be involved in transcriptional regulation.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAB domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; U37263; AAB09749.1; -.
DR HSSP; P08047; ISP2.
DR Genew; HGNC:12966; ZNF177.
DR MG; 601276; -.
DR GO; GO:0001022; P:negative regulation of transcription from P...; TAS.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 7.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 7.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 7.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 14 84 KRAB.
FT ZN_FING 124 146 C2H2-TYPE.
FT ZN_FING 152 174 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 208 230 C2H2-TYPE.
FT ZN_FING 236 258 C2H2-TYPE.
FT ZN_FING 264 286 C2H2-TYPE.
FT ZN_FING 292 314 C2H2-TYPE.
SQ SEQUENCE 321 AA; 36473 MW; E3258606C292DA17 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGQKL 12
Db 313 IHNGQKL 319

RESULT 7
GLI4_HUMAN STANDARD; PRT; 376 AA.
AC P10075; Q96CK9;
DT 01-MAR-1989 (Rel. 10, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinn finger protein GLI4 (Kruempel-related zinc finger protein 4)
DE (HKR4 protein).
GN GLI4 OR HKR4.

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22398257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Rahay J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Varra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE OF 144-272 AND 283-376 FROM N.A.
 RX MEDLINE=89096896; PubMed=2850480;
 RA Ruppert J.M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
 RA Law M.L., Seunarine H.N., O'Brien S.J., Vogelstein B.;
 RT "The GLI-Kruppel family of human genes.";
 RL Mol. Cell. Biol. 8:3104-3113(1988).
 CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -|- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -|- SIMILARITY: Contains 7 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; BC014165; AAH14165.1; -;
 DR EMBL; M20678; AAA35990.1; -;
 DR PIR; F31201; F31201.
 DR HSP; P08046; IALG.
 DR Genew; HGNC:4320; GLI4.
 DR MIN; 165280; -;
 DR GO; GO:0005634; C:nucleus; NAS.
 DR InterPro; IPR007087; Znf C2H2.
 DR Pfam; PF00096; zf-C2H2; 7.
 DR ProDom; PD000003; Znf C2H2; 4.
 DR SMART; SM00355; Znf C2H2; 7.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
 KW Zinc-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat.
 FT ZN_FING 183 205 C2H2-TYPE 1.
 FT ZN_FING 211 233 C2H2-TYPE 2.
 FT ZN_FING 239 261 C2H2-TYPE 3.
 FT ZN_FING 267 289 C2H2-TYPE 4.
 FT ZN_FING 295 317 C2H2-TYPE 5.
 FT ZN_FING 323 345 C2H2-TYPE 6.
 FT ZN_FING 351 373 C2H2-TYPE 7.
 FT CONFLICT 144 146 RVT -> ARD (IN REF. 2).
 FT CONFLICT 180 180 A -> T (IN REF. 2).
 SQ SEQUENCE 376 AA; 41145 MW; 56B1A638CC0759BC CRC64;

Query Match 59.4%; Score 38; DB 1; Length 376;
 Best Local Similarity 63.6%; Pred. No. 18;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 [1] TSPLNHNHGQK 11
 DB 255 TQHLRHNGEK 265
 RESULT 8
 FEK3 HUMAN
 ID FEK3 HUMAN STANDARD; PRT; 471 AA.
 AC Q9UK99; Q9NUX2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE F-box only protein 3.
 GN FBXO3 OR FBX3
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Oka T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 6-415 FROM N.A.
 RX MEDLINE=20003061; PubMed=10531037;
 RA Winston J.T., Koepf D.M., Zhu C., Eledge S.J., Harper J.W.;
 RT "A family of mammalian F-box proteins.";
 RL Curr. Biol. 9:1180-1182(1999).
 CC -|- FUNCTION: Probably recognizes and binds to some phosphorylated
 CC proteins and promotes their ubiquitination and degradation.
 CC -|- SUBUNIT: Part of a SCF (SKP1-cullin-F-box) protein ligase complex
 CC (By similarity).
 CC -|- SIMILARITY: Contains 1 F-box domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AK001943; BAA91991.1; -;
 DR EMBL; AF176702; AAF03702.1; -;
 DR Genew; HGNC:13582; FBXO3.
 DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; TAS.
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; TAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
 DR InterPro; IPR007474; DUF525.
 DR InterPro; IPR001810; F-box.
 DR InterPro; IPR008945; Skp1_Skp2.
 DR Pfam; PF04379; DUF525; 1.
 DR Pfam; PF00646; F-box; 1.
 DR SMART; SM00256; FBXO; 1.
 DR PROSITE; PS0181; FBXO; 1.
 KW Ub1 conjugation pathway.
 FT DOMAIN 10 56 F-BOX
 FT DOMAIN 419 452 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 453 456 POLY-ARG.
 FT DOMAIN 463 466 POLY-ARG.
 FT CONFLICT 164 164 T -> A (IN REF. 2).
 FT CONFLICT 414 415 EM -> VS (IN REF. 2).
 SQ SEQUENCE 471 AA; 54590 MW; F7AA89193E14567E CRC64;

Query Match 59.4%; Score 38; DB 1; Length 471;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGQKL 12
 Db 148 IHNGQKL 154

RESULT 9
 ID_PLO1_SCHPO STANDARD; PRT; 683 AA.

AC F50528; (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serine/threonine-protein kinase p101 (EC 2.7.1.37).
 GN PLO1 OR SPAC23C11.16
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN=972; PubMed=7744248;
 RX MEDLINE=95562899; PubMed=7744248;
 RA Okura H., Hagan I.M., Glover D.M.;
 RT "The conserved Schizosaccharomyces pombe kinase p101, required to
 RT form a bipolar spindle, the actin ring, and septum, can drive septum
 RT formation in G1 and G2 cells.";
 RL Genes Dev. 9:1059-1073(1995).
 RN [2]
 RP SEQUENCE FROM N.A.

CC Query Match 59.4%; Score 38; DB 1; Length 471;
 CC Best Local Similarity 100.0%; Pred. No. 23;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNGQKL 12
 Db 148 IHNGQKL 154

RESULT 9
 ID_PLO1_SCHPO STANDARD; PRT; 683 AA.

AC F50528; (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serine/threonine-protein kinase p101 (EC 2.7.1.37).
 GN PLO1 OR SPAC23C11.16
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN=972; PubMed=7744248;
 RX MEDLINE=95562899; PubMed=7744248;
 RA Okura H., Hagan I.M., Glover D.M.;
 RT "The conserved Schizosaccharomyces pombe kinase p101, required to
 RT form a bipolar spindle, the actin ring, and septum, can drive septum
 RT formation in G1 and G2 cells.";
 RL Genes Dev. 9:1059-1073(1995).
 RN [2]
 RP SEQUENCE FROM N.A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X85758; CAA59766.1; -;
 DR EMBL; Z98559; CAB11167.1; -;
 DR PIR; T38254; T38254.
 DR HSP; Q63450; IA06.
 DR GeneDB_Spombe; SPAC23C11.16; -;
 DR InterPro; IPR000959; POLO_box.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR008271; Ser_thr_pkin_AS.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF000659; pkinase; 1.
 DR Pfam; PF00659; POLO_box; 2.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SMO0220; S_TK; 1.
 DR PROSITE; PS00078; POLO_BOX; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
 FT DOMAIN 41 296 PROTEIN_KINASE.
 FT NP_BIND 47 55 ATP (BY SIMILARITY).
 FT BINDING 69 69 ATP (BY SIMILARITY).
 FT ACT_SITE 163 163 BY SIMILARITY.
 FT DOMAIN 500 567 POLO_BOX 1.
 FT DOMAIN 604 670 POLO_BOX 2.
 SQ SEQUENCE 683 AA; 77301 MW; F11CD0EF9B913917 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 683;
 Best Local Similarity 54.5%; Pred. No. 35;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPINTHNGQK 11
 Db 27 TPPTNLHNRKK 37

RESULT 10
 ID_S0HA_ECOLI STANDARD; PRT; 111 AA.
 AC P15373;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE HtrA suppressor protein (protein prlf).
 GN S0HA OR PRLF OR B3129.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90170878; PubMed=2407727;
 RA Baird L., Georgopoulos C.;
 RT "Identification, cloning, and characterization of the Escherichia
 RT coli s0ha gene, a suppressor of the htrA (degP) null phenotype.";
 RL J. Bacteriol. 172:1587-1594(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90094216; PubMed=2152898;
 RA Kline D.R., Phillips G.J., Silhavy T.J.;
 RT "Increased expression of the bifunctional protein prlf suppresses
 RT overproduction lethality associated with exported beta-galactosidase
 RT hybrid proteins in Escherichia coli.";
 RL J. Bacteriol. 172:185-192(1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkert G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474 (1997).
 CC -!- FUNCTION: PRF APPEARS TO BE A BIFUNCTIONAL PROTEIN, WITH THE
 CC ABILITY TO REGULATE ITS OWN EXPRESSION AS WELL AS RELIEVE THE
 CC EXPORT BLOCK IMPOSED BY HIGH-LEVEL SYNTHESIS OF THE LAMB-LACZ
 CC HYBRID PROTEIN.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M30178; AAA24638.1; -;
 DR EMBL; M32358; AAA24418.1; -;
 DR EMBL; U18997; AAA57932.1; -;
 DR EMBL; AB000394; AAC76163.1; -;
 DR PIR; A35137; A35137.
 DR EcoGene; EGI0955; soha.
 DR InterPro; IPR006339; ABrB trans reg.
 DR TIGRFAMs; TIGR01439; lp_hng_hel_AbrB; 1.
 KW Complete proteome.
 SQ SEQUENCE 111 AA; 12359 MW; 5FC0D5FF43F75D8A CRC64;

Query Match 57.8%; Score 37; DB 1; Length 111;
 Best Local Similarity 58.3%; Pred. No. 7.3;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSPLNHNGQKL 12
 DB 84 TRPFIQQGKKL 95

RESULT 11
 REFD SALTY STANDARD; PRT; 299 AA.
 ID -REFD SALTY STANDARD; PRT; 299 AA.
 AC P26392;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE dTDP-4-dehydrothamose reductase (EC 1.1.1.133) (dTDP-4-keto-L-
 DE thamose reductase) (dTDP-6-deoxy-L-mannose dehydrogenase) (dTDP-L-
 DE thamose synthetase).
 GN REFD OR STM2096.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=91260454; PubMed=1710759;
 RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Romana L.K., Reeves P.R.;
 RT "Structure and sequence of the rfb (O antigen) gene cluster of
 RT Salmonella serovar typhimurium (strain LT2).";
 RL Mol. Microbiol. 5:695-713 (1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856 (2001).
 CC -!- CATALYTIC ACTIVITY: dTDP-6-deoxy-L-mannose + NADP(+) = dTDP-4-
 CC dehydro-6-deoxy-L-mannose + NADPH.
 CC -!- PATHWAY: dTDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56793; CAA40116.1; -;
 DR EMBL; AE008792; AAL21000.1; -;
 DR PIR; S15300; S15300.
 DR PDB; 1KB2; 28-JUN-02.
 DR PDB; 1KC1; 28-JUN-02.
 DR PDB; 1KC3; 28-JUN-02.
 DR StyGene; SG10344; rfbD.
 DR InterPro; IPR005913; TDP rham reduct.
 DR Pfam; PF04321; RmlD_sub_Bind; 1.
 DR TIGRFAMs; TIGR01214; rmlD; 1.
 KW Lipopolysaccharide biosynthesis; Oxidoreductase; NADP;
 KW Complete proteome; 3D-structure.
 FT NP_BIND 7 NADP (POTENTIAL).
 SQ SEQUENCE 299 AA; 32554 MW; ABAA0476AF5ECDE7 CRC64;

Query Match 57.8%; Score 37; DB 1; Length 299;
 Best Local Similarity 58.3%; Pred. No. 22;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPLNHNGQKL 12
 DB 122 TSPLNHYGKTKL 133

RESULT 12
 GAG BIV06 STANDARD; PRT; 476 AA.
 ID GAG BIV06 STANDARD; PRT; 476 AA.
 AC P19558;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GAG polyprotein (P53) [Contains: Core proteins P17, P26, P14].
 GN GAG.
 OS Bovine immunodeficiency virus (isolate 106) (BIV).
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11658;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90223985; PubMed=2183467;
 RA Garvey K.J., Oberste M.S., Elser J.E., Braun M.J., Gonda M.A.;
 RT "Nucleotide sequence and genome organization of biologically active
 RT proviruses of the bovine immunodeficiency-like virus.";
 RL Virology 175:391-409 (1990).
 CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M32691; -; NOT ANNOTATED_CDS.
 DR HIV; M32691; GAGSBIVT06.

```

DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR008916; Retrov_Capsid_C.
DR InterPro; IPR008919; Retrov_Capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 2.
DR Core protein; Polyprotein; Zinc-finger.
FT CHAIN 1 133 MATRIX PROTEIN (P17) (POTENTIAL).
FT CHAIN 134 360 CAPSID PROTEIN (P26) (POTENTIAL).
FT CHAIN 361 476 NUCLEOCAPSID (P14) (POTENTIAL).
FT ZN_FING 403 420 CCHC-TYPE 1.
FT ZN_FING 421 438 CCHC-TYPE 2.
SQ SEQUENCE 476 AA; 53470 MW; 54D7F2B595A80269 CRC64;

Query Match 57.8%; Score 37; DB 1; Length 476;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PLNHNQK 11
DB 287 PINHQGP 295

RESULT 13
GAG_BIV27 STANDARD; PRT; 476 AA.
AC P19559;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GAG polyprotein (P53) [Contains: Core proteins P17, P26, P14].
GN GAG.
OS Bovine immunodeficiency virus (isolate 127) (BIV).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11659;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90223985; PubMed=2183467;
RA Garvey K.J., Oberste M.S., Eiser J.E., Braun M.J., Gonda M.A.;
RT "Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus.";
RL Virology 175:391-409(1990).
CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; M32690; AA91270.1; -.
DR PIR; A34742; FOLJET.
DR HIV; M32690; GAGSIV127.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR008916; Retrov_Capsid_C.
DR InterPro; IPR008919; Retrov_Capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 2.
DR Core protein; Polyprotein; Zinc-finger; Repeat.
FT CHAIN 1 133 MATRIX PROTEIN (P17) (POTENTIAL).
FT CHAIN 134 360 CAPSID PROTEIN (P26) (POTENTIAL).
FT CHAIN 361 476 NUCLEOCAPSID (P14) (POTENTIAL).
FT ZN_FING 403 420 CCHC-TYPE 1.
FT ZN_FING 421 438 CCHC-TYPE 2.

```

```

SQ SEQUENCE 476 AA; 53440 MW; FAA896BD684255FF CRC64;

Query Match 57.8%; Score 37; DB 1; Length 476;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PLNHNQK 11
DB 287 PINHQGP 295

RESULT 14
MINC_HELPJ STANDARD; PRT; 194 AA.
ID Q9ZMS1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable septum site-determining protein minc.
GN MINC OR JHP0372.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: Cell division inhibitor that blocks the formation of polar Z ring septums. Rapidly oscillates between the poles of the cell to destabilize ftsZ filaments that have formed before they mature into polar Z rings. Prevent ftsZ polymerization (By similarity).
CC -!- SUBUNIT: Interacts with minD and ftsZ (By similarity).
CC -!- SIMILARITY: Belongs to the minC family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AE001472; AAD05953.1; -.
DR PIR; H71939; H71939.
DR HAMAP; MF_00267; -.
DR InterPro; IPR005526; MinC.
DR Pfam; PF03775; MinC_C; 1.
DR Cell division; Septation; Complete proteome.
SQ SEQUENCE 194 AA; 22302 MW; 33093F48637D0FA8 CRC64;

Query Match 56.2%; Score 36; DB 1; Length 194;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 NIHNQKL 12
DB 117 NIHNQAKI 124

RESULT 15
MINC_HELPJ STANDARD; PRT; 195 AA.
ID MINC_HELPJ
AC O25693;
DT 16-OCT-2001 (Rel. 40, Created)

```

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable septum site-determining protein minC.
GN MINC OR HP1053.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP
SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=9739467; PubMed=9252185;
RA Tomb J.-P., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Cocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547 (1997).
CC -!- FUNCTION: Cell division inhibitor that blocks the formation of
CC polar Z ring septums. Rapidly oscillates between the poles of the
CC cell to destabilize ftsZ filaments that have formed before they
CC mature into polar Z rings. Prevent ftsZ polymerization (By
CC similarity).
CC -!- SUBUNIT: Interacts with mind and ftsZ (By similarity).
CC -!- SIMILARITY: Belongs to the minC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000613; AAD08105.1; ALT_INIT.
DR TIGR; HP1053; -.
DR HAMAP; MF_00267; -.
DR InterPro; IPR005526; MinC.
DR Pfam; PF03775; MinC; 1.
KW Cell division; Septation; Complete proteome.
SQ SEQUENCE 195 AA; 22372 MW; 5C2DE7235B33C77B CRC64;
-----
Query Match 56.2%; Score 36; DB 1; Length 195;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 NIHNGQKL 12
Db 118 NIHNGAKI 125
| | | | |
| | | | |

```

Search completed: September 15, 2004, 09:57:17
Job time : 25 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: September 15, 2004, 09:46:35 ; Search time 115 Seconds
(without alignments)
32.924 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 64
Sequence: 1 TSPLNTHNGQKL 12
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				SPTRMBL 25:*			
				1: sp_archaea:*			
				2: sp_bacteria:*			
				3: sp_fungi:*			
				4: sp_human:*			
				5: sp_invertebrate:*			
				6: sp_mammal:*			
				7: sp_mmc:*			
				8: sp_organelle:*			
				9: sp_phage:*			
				10: sp_plant:*			
				11: sp_rodent:*			
				12: sp_virus:*			
				13: sp_vertebrate:*			
				14: sp_unclassified:*			
				15: sp_rvirus:*			
				16: sp_bacteriaph:*			
				17: sp_archaeap:*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	67.2	321	4 Q96ER2	Q96ER2 homo sapien
2	42	65.6	1302	5 Q8MRC4	Q8MRC4 drosophila
3	42	65.6	1343	16 Q8XAY4	Q8XAY4 escherichia
4	42	65.6	2531	5 O16004	O16004 lytechinus
5	41	64.1	392	16 Q893D8	Q893D8 clostridium
6	41	64.1	427	10 Q9SZV7	Q9SZV7 arabidopsis
7	40	62.5	331	3 O14097	O14097 schizosacch
8	40	62.5	334	5 Q9GUZ8	Q9GUZ8 aikopleura
9	40	62.5	438	16 Q88DM7	Q88DM7 pseudomonas
10	40	62.5	630	10 Q8LXK6	Q8LXK6 oryza sativ
11	39	60.9	243	16 Q8ZHG4	Q8ZHG4 yersinia pe
12	39	60.9	243	16 Q8EIK7	Q8EIK7 shewanella
13	39	60.9	392	4 Q9H807	Q9H807 homo sapien
14	39	60.9	563	16 Q9X8S9	Q9X8S9 streptomyce
15	39	60.9	738	4 Q9ULAI	Q9ULAI homo sapien
16	39	60.9	855	17 Q8TT99	Q8TT99 methanosarc

17	39	60.9	899	5 Q8IEN8	Q8IEN8 plasmodium
18	39	60.9	914	12 Q9IF30	Q9IF30 bovine aden
19	39	60.9	938	12 Q9W8S1	Q9W8S1 porcine ade
20	39	60.9	939	12 Q84178	Q84178 porcine ade
21	39	60.9	970	5 Q96115	Q96115 drosophila
22	39	60.9	1463	5 Q86NZ0	Q86NZ0 drosophila
23	39	60.9	2243	5 Q9VGP1	Q9VGP1 drosophila
24	38	59.4	150	16 Q9CJB1	Q9CJB1 lactococcus
25	38	59.4	173	4 Q9UKC5	Q9UKC5 homo sapien
26	38	59.4	188	2 Q93MT1	Q93MT1 photorhabdu
27	38	59.4	277	2 Q8VTA6	Q8VTA6 synechococc
28	38	59.4	324	5 Q22161	Q22161 caenorhabdi
29	38	59.4	367	11 Q91VI9	Q91VI9 mus musculu
30	38	59.4	406	16 Q8ZMW4	Q8ZMW4 salmonella
31	38	59.4	406	16 Q8Z4I6	Q8Z4I6 salmonella
32	38	59.4	415	11 Q9JIE4	Q9JIE4 mus musculu
33	38	59.4	444	10 P93716	P93716 petunia hyb
34	38	59.4	471	4 Q9H0V2	Q9H0V2 homo sapien
35	38	59.4	472	16 Q86X90	Q86X90 homo sapien
36	38	59.4	472	16 Q8CF59	Q8CF59 staphylococ
37	38	59.4	480	11 Q9D2J5	Q9D2J5 mus musculu
38	38	59.4	480	11 Q9DC63	Q9DC63 mus musculu
39	38	59.4	480	11 Q8CEJ0	Q8CEJ0 mus musculu
40	38	59.4	480	11 Q8C7I0	Q8C7I0 mus musculu
41	38	59.4	627	5 Q18586	Q18586 caenorhabdi
42	38	59.4	693	10 Q9SKH6	Q9SKH6 arabidopsis
43	37.5	58.6	2588	16 Q8CVE4	Q8CVE4 leptospira
44	37	57.8	111	16 Q8XAE1	Q8XAE1 escherichia
45	37	57.8	112	4 Q9H4Y3	Q9H4Y3 homo sapien

ALIGNMENTS

RESULT 1

Q96ER2	PRELIMINARY;	PRT;	321 AA.
ID	Q96ER2		
AC	Q96ER2		
DT	01-DEC-2001 (TRENBLrel. 19, Created)		
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)		
DE	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
ON	NCBI_TaxID=9606;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis, and Embryonic carcinoma;		
RA	Strausberg R.; 2001 to the EMBL/GenBank/DBJ databases.		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
DR	EMBL; BC012012; AAH12012.1; -		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR001909; KRAB.		
DR	InterPro; IPR007087; Znf_C2H2.		
DR	InterPro; IPR007086; Znf_C2H2_sub.		
DR	Pfam; PF01352; KRAB; 1.		
DR	Pfam; PF00096; zf-C2H2; 7.		
DR	PRINTS; PR00048; ZINC_FINGER.		
DR	ProDom; PD000003; Znf_C2H2; 6.		
DR	SMART; SM00349; KRAB_1.		
DR	SMART; SM00355; Znf_C2H2; 7.		
DR	PROSITE; PS00805; KRAB; 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.		
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.		
KW	Hypothetical protein; Metal-binding; Nuclear protein; Zinc;		
KW	Zinc-finger.		
SQ	SEQUENCE 321 AA; 36369 MW; 62BE342C8B7067D9 CRC64;		

Query Match 67.2%; Score 43; DB 4; Length 321;

```

Best Local Similarity 80.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 PLNIHQKQL 12
Db 310 PKSIHQKQL 319

RESULT 2
Q8MRC4 PRELIMINARY; PRT; 1302 AA.
AC Q8MRC4;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE R214657P.
GN MDR65 OR CG10181.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celnikier S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY121669; AAM51996.1; -.
DR FlyBase; FBgn0004513; Mdr65.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP-binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding.
SQ SEQUENCE 1302 AA; 143813 MW; ECDA47C2F3F4DC46 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 1302;
Best Local Similarity 72.7%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSLPNIHQKQL 11
Db 673 TSLPNIHQKQL 683

RESULT 3
Q8XAY4 PRELIMINARY; PRT; 1343 AA.
AC Q8XAY4;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Orf, hypothetical protein.
GN YDEK OR Z2195 OR ECS2117.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=83334;

```

```

[1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Horda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005354; AGS6256.1; -.
DR EMBL; AP002557; BAB35540.1; -.
DR PIR; D85724; D85724.
DR PIR; E90893; E90893.
DR InterPro; IPR000437; Prok_lipo_prot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 1343 AA; 138568 MW; 4C2456739907B63D CRC64;

Query Match 65.6%; Score 42; DB 16; Length 1343;
Best Local Similarity 88.9%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHQKQL 12
Db 991 LNIHQKQL 999

RESULT 4
O16004 PRELIMINARY; PRT; 2531 AA.
ID O16004;
AC O16004;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Notch homolog.
OS Lytechinus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinzoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopneustidae;
OC Lytechinus.
CX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97454256; PubMed=9310331;
RA Sherwood D.R., McClay D.R.;
RT "Identification and localization of a sea urchin Notch homologue:
RT insights into vegetal plate regionalization and Notch receptor
RT regulation."
RL Development 124:3363-3374(1997).
DR EMBL; AF000634; AAB82088.1; -.
DR HSSP; P01132; 1EGF.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.

```

```

DR InterPro; IPR008297; Notch.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00084; ANK_REPEAT; 5.
DR PROSITE; PS02937; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 21.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS01187; EGF_CA; 20.
DR PIRSP; PIRSF002279; Notch; 1.
KW ANK repeat; EGF-like domain; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 2531;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLEINHQQ 10
Db |||:|||||
2097 TSPMDVHNGE 2106

RESULT 5
Q893D8 PRELIMINARY; PRT; 392 AA.
ID Q893D8;
AC Q893D8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79).
GN CTO1889
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wieser A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR EMBL; AE015942; AAC36404.1; -.
DR GO; GO:0008825; F:cyclopropane-fatty-acyl-phospholipid syntha...; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008610; P:lipid biosynthesis; IEA.
DR InterPro; IPR003333; CMAW.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF02353; CMAW; 1.
DR Methyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 392 AA; 45910 MW; 4E4D927C532C5037 CRC64;

Query Match 64.1%; Score 41; DB 16; Length 392;
Best Local Similarity 77.8%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNHGQKL 12
Db |||:|||||
160 LNLHSGQKL 168

RESULT 6
Q9SZV7 PRELIMINARY; PRT; 427 AA.
ID Q9SZV7;
AC Q9SZV7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Hypothetical protein.
DE PEG3.70 OR ATG30040.
GN PEG3.70 OR ATG30040.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Arabidopsids sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC Arabidopsids sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC Arabidopsids sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC Arabidopsids sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC Arabidopsids sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078464; CAB43839.1; -.
DR EMBL; AL161576; CAB80997.1; -.
DR PIR; T08980; T08980.
DR GO; GO:0004194; F:pepsin A activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR004461; Peptidase_A1.
DR InterPro; IPR009007; Pept_A_acid.
DR PRINTS; PR00792; PEP SIN.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47357 MW; F6F42BDD938B3225 CRC64;

Query Match 64.1%; Score 41; DB 10; Length 427;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TSPLEINHNG 9
Db |||:|||||
263 TTPLEIHNG 271

RESULT 7
O14097 PRELIMINARY; PRT; 331 AA.
ID O14097;
AC O14097;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE Proline-rich protein, predicted coiled-coil region, WW domain.
GN SPAC2F3.14C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Oliver K., Harris D., Barrall B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z59165; CAB16268.1; -.
DR PIR; T38546; T38546.
DR GeneDB.SPombe; SPAC2F3.14C; -.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00397; WW; 1.

```

```

SQ SEQUENCE 331 AA; 37669 MW; EA888EACCC4CBB7B1 CRC64;
Query Match 62.5%; Score 40; DB 3; Length 331;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPLNHHNGQKL 12
DB 271 SSLNSHNGQSL 281

RESULT 8
Q9GU28 PRELIMINARY; PRT; 334 AA.
AC Q9GU28;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Brachyury protein.
GN CDIT.
OS Oikopleura dioica.
OC Eukaryota; Metazoa; Chordata; Urochordata; Appendicularia;
OC Oikopleuridae; Oikopleura.
OX NCBI_TaxID=34765;
RN [1]
SEQUENCE FROM N.A.
RA Bassham S., Postlethwait J.H.;
RT "Brachyury (T) expression in embryos of a larvacean urochordate,
RT Oikopleura dioica, and the ancestral role of brachyury."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF204208; AAG22592.1; -
DR HSP; P24781; 1XBR.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR008967; P53-like.
DR InterPro; IPR001699; TF-T-box.
DR Pfam; PF00907; T-box; 1.
DR PRINTS; PR00937; TBOX.
DR SMART; SM00425; TBOX. 1.
DR PROSITE; PS01283; TBOX_1; 1.
DR PROSITE; PS0252; TBOX_3; 1.
DR PROSITE; PS0252; TBOX_3; 1.
SQ SEQUENCE 334 AA; 37574 MW; 02AEA233B5800F50 CRC64;

Query Match 62.5%; Score 40; DB 5; Length 334;
Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TSPLNHHNGQKL 12
DB 125 TSKLHHNGQQL 136

RESULT 9
Q88DM7 PRELIMINARY; PRT; 438 AA.
AC Q88DM7;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Membrane-bound lytic murein transglycosylase, putative.
GN PP4798.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weiner C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,

```

```

RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016792; AAN70367.1; -
DR TIGR; PP4798; -
DR GO; GO:0000270; P:peptidoglycan metabolism; IEA.
DR InterPro; IPR002477; PG-binding.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF01471; PG-binding_1; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 438 AA; 48000 MW; 615552F281AAF1BC CRC64;

Query Match 62.5%; Score 40; DB 16; Length 438;
Best Local Similarity 63.6%; Pred. No. 48;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SPLNHHNGQKL 12
DB 123 SPLRVNGKKL 133

RESULT 10
Q8LNNK6 PRELIMINARY; PRT; 630 AA.
ID Q8LNNK6 PRELIMINARY; PRT; 630 AA.
AC Q8LNNK6;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OGJNBA0071120.12.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA McCombie W.R., Spiegel L., de la Bastide M., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zutavern T., Ballja V., Bell M., Baker J.,
RA Miller B., Katzenberger F., Muller S., King L., Sullivan P., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OGJNBA0071120, from chromosome 10, complete sequence."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RC The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10."
RL Science 300:1566-1569(2003).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RC Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC074355; AAM74250.1; -
DR EMBL; AE017062; AAP52344.1; -
DR Gramene; Q8LNNK6; -
KW Hypothetical protein.
SQ SEQUENCE 630 AA; 67623 MW; AD2A264C95EE8412 CRC64;

Query Match 62.5%; Score 40; DB 10; Length 630;
Best Local Similarity 72.7%; Pred. No. 70;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TSPLNHNQK 11
DB 469 TSPLAQNGKK 479

RESULT 11
ID Q82HG4 PRELIMINARY; PRT; 243 AA.
AC Q82HG4; 2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein YPO0934.
GN YPO0934 OR Y3320.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaiha M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RT Nature 413:523-527(2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Bourin A., Mayhew G.P., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Landler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AJ414145; CAC89777.1; -
DR EMBL; AF01114; AF01114.
DR InterPro; IPR004382; Cons_hypoth46.
DR Pfam; PF04452; DUF558; 1.
DR TIGRFAMs; TIGR00046; TIGR00046; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 26920 MW; BFF54E709F9A37CC CRC64;

Query Match 60.9%; Score 39; DB 16; Length 243;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNHNQK 10
DB 76 SPLNLHLQ 84

RESULT 12
ID Q8EIK7 PRELIMINARY; PRT; 243 AA.
AC Q8EIK7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein TIGR00046.
GN SO832.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.

```

```

OX NCBI_TaxID=70863;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Barry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uitterlind L.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015528; AAN53908.1; -
DR TIGR; SO832; -
DR InterPro; IPR004382; Cons_hypoth46.
DR InterPro; IPR006700; DUF558.
DR Pfam; PF04452; DUF558; 1.
DR TIGRFAMs; TIGR00046; TIGR00046; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 26601 MW; 1713957210A6286B CRC64;

Query Match 60.9%; Score 39; DB 16; Length 243;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPLNHNQK 10
DB 76 SPLNLHLQ 84

RESULT 13
Q8H07
ID Q8H07 PRELIMINARY; PRT; 392 AA.
AC Q8H07;
DT 01-NAR-2001 (TReMBLrel. 16, Created)
DT 01-NAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ14011.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi N., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saio K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AK024073; BAB14816.1; -
DR HSSP; P08046; IAIH.
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 12.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 7.
DR SMART; SM00355; Znf_C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 12.
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW Zinc-finger.
SQ SEQUENCE 392 AA; 45453 MW; 33984A6D50082D2D CRC64;

```

```

Query Match          60.9%; Score 39; DB 4; Length 392;
Best Local Similarity 77.8%; Pred. No. 66;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 4 LNIHNGOKL 12
Db 159 LRIHNGEKL 167

RESULT 14
Q9X8S9 PRELIMINARY; PRT; 563 AA.
AC Q9X8S9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein SC03893.
GN SC03893 OR SCH24.15C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Harris D.E., Quail M.A., Kieser H.,
RA Thomson N.R., James K.D., Harris D.E., Chandra G., Chen C.W., Collins M.,
RA Harper D., Bateman A., Brown S., Hidalgo J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Goble A., Goble A., Hildalgo J., Oliver K., O'Neill S.,
RA Huang C.H., Kieser T., Larke J., Murphy J., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AJ939118; CAB42719.1; -.
DR PIR; T36580;
DR InterPro; IPR008979; Gal_bind_like.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 563 AA; 60226 MW; 7FB4DDP8B5E4EB72 CRC64;

Query Match          60.9%; Score 39; DB 16; Length 563;
Best Local Similarity 58.3%; Pred. No. 96;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TSPPLNIHNGOKL 12
Db 57 TSPPLHSGHKL 68

RESULT 15
Q9ULAI PRELIMINARY; PRT; 738 AA.
AC Q9ULAI;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R31155.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,

```

```

RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 2.3 Mb region containing a zinc finger (ZNF)
RT cluster in 19q13.4.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC013256; AAF06067.1; -.
DR HSSP; P08046; 1A1J.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 15.
DR ProDom; PD000003; Znf_C2H2; 7.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 15.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 15.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 738 AA; 84252 MW; 9E158059D5C18527 CRC64;

Query Match          60.9%; Score 39; DB 4; Length 738;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LNIHNGOKL 12
Db 505 LRIHNGEKL 513

Search completed: September 15, 2004, 09:59:19
Job time : 118 secs

```